University of Zagreb

Faculty of Science

Department of Biology

Doris Zorić

Characterization of dispersed alpha satellite DNA elements in human genome

Graduation Thesis

Zagreb, 2018

This thesis was made at Ruđer Bošković Institute, under supervision of Dr. Đurđica Ugarković, Assoc. Prof. and Dr. Isidoro Feliciello from University of Naples Federico II, and submitted for evaluation to Department of Biology, Faculty of Science in order to acquire title the Master of molecular biology.

BASIC DOCUMENTATION CARD

University of Zagreb

Faculty of Science

Department of Biology Graduation Thesis

CHARACTERIZATION OF DISPERSED ALPHA SATELLITE DNA ELEMENTS IN HUMAN GENOME

Doris Zorić

Roosevelt Square 6, 10000 Zagreb, Croatia

Satellite DNAs are tandemly repeated sequences present in long uninterupted arrays in genetically silenced heterochromatic regions. Genome sequencing projects revealed the presence and the distribution of repetitive satellite DNA in the vicinity of genes, or in other words, the dispersion of satellite DNA elements across the euchromatic portion of genome. The presence of functional elements such as promoters and transcription factors binding sites within a number of satelites led to assumption that satellite DNAs may act as gene expression *cis*-regulatory elements. The main goal of this research is to determine the abundance and the distribution of dispersed elements of alpha satellite DNA in human genome and to propose them as genetic polymorphisms in the population, as well as to investigate the potential regulatory role of those elements by DNA methylation analysis before and after the heat-shock treatment.

(51 pages, 16 figures, 11 tables, 35 references, original in: English)

Thesis deposited in the Central Biological Library

Key words*:* alpha satellite DNA, *cis*-regulatory elements, insertion polymorphism, adaptation to heat stress

Supervisor: Dr. Đurđica Ugarković, Assoc. Prof.

Reviewers: Dr. Đurđica Ugarković, Assoc. Prof.; Dr. Zlatko Liber, Prof.; Dr. Goran Kovačević, Assoc. Prof.

Thesis accepted: 1. 2. 2018.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek Diplomski rad

KARAKTERIZACIJA RASPRŠENIH ELEMENATA ALFA-SATELITNE DNA U LJUDSKOM GENOMU

Doris Zorić

Rooseveltov trg 6, 10000 Zagreb, Croatia

Satelitne DNA su uzastopno ponavljajuće sekvence koje su prisutne u dugačkim neprekinutim nizovima u genetički utišanim heterokromatinskim regijama. Projekti sekvenciranja genoma pokazali su prisutnost i distribuciju ponavljanja satelitne DNA u eukromatinskom dijelu genoma u blizini samih gena, odnosno raspršenost elemenata satelitne DNA u genomu. Prisutnost funkcionalnih elemenata poput promotora i mjesta vezanja transkripcijskih faktora unutar nekih satelita doveli su do pretpostavke da satelitne DNA mogu djelovati kao *cis*-regulatorni elementi ekspresije gena. Cilj istraživanja jest utvrditi zastupljenost i distribuciju raspršenih elemenata alfa-satelitne DNA u ljudskom genomu, uočiti moguće polimorfizme u populaciji te ispitati regulatornu ulogu tih elemenata analizom metilacije DNA prije i nakon temperaturnog stresa.

(51 stranica, 16 slika, 11 tablica, 35 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: alfa-satelitna DNA, *cis*-regulatorni elementi, insercijski polimorfizam, prilagodba na temperaturni stres

Mentor: Dr. sc. Đurđica Ugarković, izv. prof.

Ocjenitelji: Dr. sc. Đurđica Ugarković, izv.prof.; Dr. sc. Zlatko Liber, red. prof.; Dr. sc. Goran Kovačević, izv. prof.

Rad prihvaćen: 1. 2. 2018.

Contents

List of abbreviations

5-hmC – 5-hidroxy-methyl-cytosine

5-mC – 5-methyl-cytosine

ANOVA – one-way analysis of variance

ATP – adenosine triphosphate

bp – base pair

Cas9 - CRISPR-associated protein-9 nuclease

CENH3 – centromeric histone H3 variant

CENP – centromere protein

CENP-C– centromere protein C

ChIP – chromatin immunoprecipitation

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

df – degree of freedom

DNA – deoxyribonucleic acid

dNTP – deoxynucleotide triphosphate

dsRNA – double-stranded RNA

EDTA - ethylenediaminetetraacetic acid

et al – *et alii* (lat.); „and others“

etc. – *et cetera* (lat.); „and the rest (of the things)“

H3K9 – lysine 9 of histone H3

H3K9me3/2 - tri-/di-methylation at lysine 9 of histone H3

hnRNP – heterogeneous nuclear ribonucleoprotein

HORs – high order repeats

HP1 – heterochromatin protein 1

HSF1 – heat shock factor 1

hTERT – human telomerase reverse transcriptase

ID – identitiy document

MHIIβ – Major Histocompatibility Complex II Beta

MS – mean square

NCBI - National Center for Biotechnology Information

NR3C1 - Nuclear Receptor Subfamily 3 Group C Member 1

PBL – peripheral blood leukocytes

PCR – polymerase chain reaction

PEV – position effect variegation

RDRC – RNA-directed RNA polymerase complex

RISC – RNA-induced silencing complex

RITS – RNA-induced transcriptional silencing complex

RNA – ribonucleic acid

RNA pol II – RNA-polymerase II

Sam 68 - Src-associated substrate in mitosis of 68 kDa

SF – suprachromosomal families

SF2/ASF – pre-mRNA-splicing factor SF / alternative splicing factor

siRNA – small interference ribonucleic acid

SNP – single-nucleotide polymorphism

SRp30c – serine and arginine rich pre-mRNA splicing factor

TAE – Tris-acetate-EDTA

TCAST – *Tribolium castaneum* satellite DNA

TE – Tris-EDTA

unspec. – unspecific

ZNF675 – zinc finger protein 675

1. Introduction

1.1. Satellite DNAs

Satellite DNAs are tandemly repeated sequences that are present as long uninterrupted arrays, predominantly in genetically silent heterocromatic regions. Basic repeat units of satellite DNAs are structured as complex sequences and organized in distinct/charachteristic/defined monomers. For instance, the basic repetitive element of human alpha (α)-satellite DNA is 171-bp monomer and it represents a main structural element of centromeric and pericentromeric regions. Furthermore, the most abundant mouse pericentromeric gama (γ)-satellite is composed of 234 bp monomers that puts it in the same group with other complex repetitive units. On the contrary, satellite DNAs might as well be composed out of short simple repeats, such as 5 bp monomers of human alpha satellite 3 (Ugarković, 2005).

Human alpha satellite DNA comprises 3-5% of each chromosome. Alpha satellite DNA was originally isolated as a highly repetitive component of the African green monkey genome (Maio, 1971). Homologous alpha repeats were subsequently identified in all primates including humans at the centromeric and pericentromeric regions of all chromosomes. Tandemly repeated 169 - 171 bp AT-rich monomers are organised in head-to-tail orientation (Figure 1). Unlike in African green monkey, where alpha satellite DNA exhibits simple organization, in other primates alpha satellite monomers can exhibit mutual divergence up to 50%. Hierarchicaly organized monomers create complex repeating units called high order repeats (HORs). Those units are composed of monomers that diverge in lenght from 2 to over 30 (Ugarković, 2013).

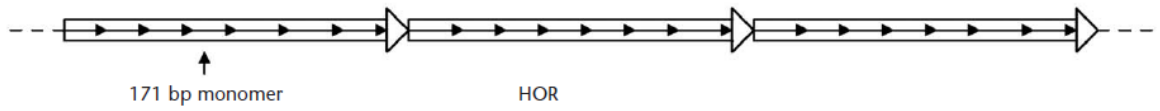


Figure 1. Organization of alpha satellite DNA. 171 bp long monomers are organized in head-to-tail fashion, thus forming complex repeat units named HORs (taken from Ugarković, 2013).

Eukaryotic genome is usually encompassed with several satellite DNAs. Subfamilies of satellite DNAs or different satellites can coexist within a genome forming a so called satellite library (Ugarković, 2013). For example, in beetle *Tribolium castaneum*, two major abundant satellite subfamilies Tcast1a and Tcast1b together comprise 35% of genomic DNA (Feliciello et al, 2011). The fundamental unit of human alpha satellite DNA is present in 12 distnct variants of diverged 171 bp monomers, which originated from 2 or 3 different ancestral monomers. Monomer types are grouped in five different alphoid „suprachromosomal families" (SF), which are charachterized by a unique set of monomeric types that alternate to form the ancestral repeat unit. Usually, alpha satellite DNA of one chromosome belongs to a single SF, but some chromosomes contain alpha satellite DNA belonging to more than one SF. Since HORs that are tandemly repeated within one long satellite array exhibit low sequence divergence (up to 5 %) and divergence between chromosomes reaches 20 %, HORs are considered to be chromosome specific (Ugarković, 2013).

Rapid horizontal spread of newly occuring mutations amongst monomers within a genome is enabled by recombinational mechanisms such as gene conversion and unequal crossing-over. These mechanisms lead to homogenisation of changes amongst repeats and their subsequent fixation in reproductive population through a process called molecular drive. This mode of horizontal evolution is known as concerted evolution, and is charachteristic for repetitive families. Apart from its influence in horizontal spread of mutations, unequal crossing-over also affects the lenght of satellite arrays through changes of satellite DNA copy number. These changes may also rise from replication slippage, rolling circle replication or conversion-like mechanisms through a short evolutionary period. All these mechanisms as a result have a high turnover of this part of the eukaryotic genome (Ugarković, 2013). Therefore, satellite DNAs exhibit divergence in sequence and copy number even between closely related species (Ugarković & Plohl, 2002). Comparison of centromeric sequences between different primate species reveals a series of amplification events resulting in replacing the „old" and spreading of „new" alpha satelite subfamilies, thus representing the mechanism by which alpha satellite evolves (Ugarković, 2013).

1.2. The role of satellite DNAs in the establishment of heterochromatin

The genomic material in eukaryotic nuclei can be roughly partitioned in two cytologically distinct entitites termed euchromatin and heterochromatin. Heterochromatin is originally defined as a portion of the genome that remains condensed through-out the cell cycle and it is generally associated with telomers and pericentromeric regions, thus representing a significant fraction of most eukaryotic genomes (Rizzi et al, 2004).

In addition to the maintainance of chromosomal integrity and stability, heterochromatin structures are involved in the inactivaion of genes normally resident in euchromatic domains. This was suggested by two key observations: the inactivation of X chromosome in mammals, a mechanism that leaves the inactive X as a visibly stained Barr body, and the possition effect variegation (PEV) in fruit fly *Drosophilla melanogaster*, a mosaic pattern of expression due to transposition of genes near heterochromatin regions. Heterochromatic regions consist predominatly of repetitive DNA, including satellite sequences and middle repetitive sequences related to transposable elements and retroviruses, and they are typically gene poor (Rizzi et al, 2004).

Satellite DNAs are the main constituents of heterochromatin. Despite their structural divergence, they associate strongly with several proteins and act as a centromere-building element in the formation of unique centromeric heterochromatin (Heinkoff & Dalal, 2005).

Many different satellite DNAs are composed of A or T clusters followed by the regular phasing of three or more A+T tracts combined with the presence of dyad structures. Periodic distribution of AT tracts produces a significant curvature of DNA helix axis, resulting in superhelical tertiary structure important for tight packing of DNA into heterochromatin (Ugarković, 2005).

The formation of dyad structures could potentially be acomplished through palindromic sequences, which are common elements of centromeric and pericentromeric satellite DNAs in budding yeast, insects and human. Particularly, some palindromic sequences could be recognized by DNA-binding proteins such as transcription factros. For example, homeodomain protein Pax3, known for its important role during neurogenesis, binds short palindromes present within the major mouse satellite DNA (pers. Commun. T. Jenuwein). Also, it has been revealed that topoisomerase II recognizes and cleaves a specific hairpin structure formed by alpha satellite DNA. The importance of this finding consists in possible cohesion role of the subpopulation of topoisomerase II at centromeres (Pezer, 2012).

In addition to conserved motifs, many variable regions of satellite DNAs may also be functionaly important due to their interaction with rapidly evolving proteins, such as centromere-specific histone variant – CENH3, which replaces histone H3 in centromeric nucleosomes and it is required for the proper chromosome distribution during cell division (Ugarković, 2005).

However, satellite DNAs are not considered a prerequisite for the centromere establishment. Instead, they are proposed to drive the adaptive evolution of specific centromeric histones (Cooper, Heinkof, 2004). Recent data indicate that the evolution of satellite DNA is as well influenced by the selective constraint related to interactions of satellite DNA with the specific proteins necessary for heterochromatin formation and its role in the control of gene expression (Ugarković, 2005).

1.3. Hypothesised mechanisms of amplification and distribution of satellite DNA through genome

Amplification of a satellite sequence is reported to occur as a result of unequal crossing-over or a duplicative transposition, and it appears to be a random process without any correlation with phylogeny of the species. After the discovery of human extrachromosomal elements originating from satellite DNA, alternative mechanisms of amplification have been suggested. Precisely, it has been proposed that satellite sequences excised from their chromosomal loci via intrastrand recombination are amplified by rolling-circle replication, followed by reintegration of tandem arrays into the genome (Felicielo et al, 2006). Such reintegration, however, might as well be guided by targeting genomic sequence based on its homology with short streches of TCAST satellite DNA extrachromosomal circles. Integrated TCAST sequences are mainly composed of interspread elements belonging to subfamilies Tcast1a and Tcast1b, which indicates that dispersed euchromatic TCAST elements formed by duplication of heterochromatin copies (Brajković et al, 2012).

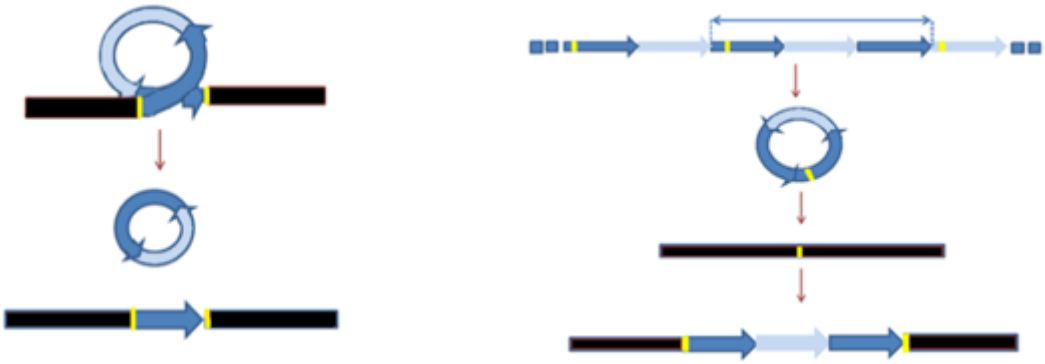


Figure 2. The proposed model of generation of dispersed satellite-like elements. Satellite sequences excised from their chromosomal loci via intrastrand recombination are amplified by rolling-circle replication, followed by reintegration of tandem arrays into the genome (left). The reintegration can be guided by targeting genomic sequence based on its homology with satellite DNA extrachromosomal circles (right) (taken and modified from Akrap, 2013).

1.4. Transcription of satellite DNAs

Transcripts of satellite DNAs have been reported in several organisms including vertebrates, invertebrates and plants (*Mus musculus*, *Drosophila melanogaster*, ...). In most species, satellite DNAs are transcribed temporally during certain developmental stages (as reported in newts *Triturus cristatus carniflex* and *Notophtalamus viridescens*) or differentially expressed in tissues, organs or cell types (such as neural crest, the head and the heart during early embriogenesis in chick and zebrafish) (Ugarković, 2005). In mammals, accumulation of centromeric and pericentromeric transcripts occurs at transcriptional level in the course of proliferation and cell cycle, differentiatioon of myoblasts, in heat shocked cells and in cancer cells, and it is mediated by RNA Polymerase II (Pezer et al, 2012).

In broad sense, there are two distinct types of transcript: one corresponding to satellite-repeat RNA and the other containing an satellite-like sequence in its 5'- and 3'- untranslated regions. Most of them are present as polyadenylated RNA in cytoplasm, but some of them are found exclusivley in nucleus. For instance, there is a strong association of satelite DNA transcripts with Y-chromosome loops in primary spermatocyte nuclei of *D. melanogaster*. In some cases, transcription is strand-specific (mouse) or it proceedes from both DNA strands (Hymenoptera) (Ugarković, 2005).

Given their relatively simple sequence and lack of any significant open frame, transcription of satellite DNA was at first ascribed to read-through from upstream genes and

transposable elements. However, promoter elements, transcription start sites and binding motifs for transcription sites have been mapped within some satellites. For example, promoters have been reported in wasp *Diadromus pulchellus*, in schisostome satellite DNA, in distant families of salamanders, in beetle species *Palorus ratzeburgii* and *Palorus subdepressus* etc. In *Drosophila*, GAGA transcription factor was found associated with heterochromatin throughout the cell cycle by interaction with a GC/AT-rich subset of satellite DNA repeats (Pezer et al, 2012).

Active transcription of centromeric alpha satellite DNA was shown to be crucial for centromere/kinetochore assembly and function during cell division (Bergmann et al, 2012). The transcription at human centromere proceeds in the form of long, single-stranded alpha satellite DNA transcripts encompassing a few satellite monomers. In addition to being functional components of the kinetochore, the transcripts are also required for the association of kinetochore proteins CENP-C1 and inner CENP in the human interphase nucleolus as well as at the centromere. Maintenance of the correct level of centromeric transcripts was shown to be indespensible for the proper cell division (Hall et al, 2012). Decrease in centromere transcription affects CENP-C deposition and is concomitant with an increase in lagging chromosomes during cell division (Chan et al, 2012). However, overexpression of alpha satellite sequence can lead to genome instability and oncogenesis (Ting et al, 2011).

Variety of examples indicates the regulatory role of satellite DNA transcripts, although in many cases it is still elusive and hypothetical. Regulatory interactions can target RNAs, DNA or proteins, or may as well involve secondary or tertiary RNA structures and RNA-mediated catalysis (Ugarković, 2005). Some transcripts have the ability to adopt hammerhead-like secondary structures and function as ribozymes with self-cleavage activity, either in *cis*- or in *trans*-. The hammerhead ribozyme structures associated with transcribed satellite DNA sequences have been found in salamanders, schistostomes and Dolichopoda save crickets. So far, all of them have been shown to self-cleave in *cis*, but the physiological role of ribozymes is yet unknown (Pezer et al, 2012).

1.5. The role of satellite DNA transcripts in thermoregulation

Stress treatments drastically affect the structure and the function of eukaryotic cell. Heat shock is one of the possible stress treatments that triggers numerous effects (Rizzi et al, 2004).

Recently, it has been shown that heat shock induces transcription of a subset of human pericentromeric satellite III. Human satellite III, which is specifically expressed under stress conditions, contains a binding motif for the heat-shock transcription factor 1 that drives RNA polymerase II transcription. The stress response consists in the recruitment of splicing factors by long, single-stranded polyadenylated satellite II transcripts to nuclear stress granules, thus sequestering splicing factors so they can regulate splicing during stress (Ugarković, 2005). Exclusively detectable in human cells, stress bodies consist of clusters of perichromatin granules and act as sites of accumulation of heat shock factor 1 (HSF1) and a subset of RNA processing factors, including components of heterogeneous nuclear ribonucleoprotein complexes (hnRNP HAP and hnRNP M), splicing factors SF2/ASF and SRp30c and RNA-binding protein Sam 68. As suggested by their colocalization with stress bodies, pericentromeric heterochromatic regions on human chromosomes 9, 12 and 15 act as a recruiting centers. Moreover, q1.2 band of chromosome 9 contains satellite III elements with sequence motifs that are similar to HSF1 binding site. Consequently, the interaction between satellite III and HSF1 is proposed to direct the recruitment of HSF1 to stress bodies.

Further association between satellite DNA and transcription factors has been reported in mice, where transcription factor YY1 associates with pericentromeric γ-satellite DNA. YY1 bellongs to Polycomb group of proteins involved in downregulation of homeotic genes and represents a link between two silencing states (Ugarković, 2005).

A setup of frequently changing number of repeats allows faster evolution, thereby promoting the adaptation to the changing environment. The discovery of short satellite segments interspersed among the genes suggest a possible regulatory role of these sequences. For instance, 32 bp- satellite repeats located in intron of the Major Histocompatibility Complex II gene (MHIIβ) regulate MHIIβ gene expression in fish *Salvelinus fontinalis* based on the temperature and the number of satellite repeats. This example represents an evidence that temperature-sensitive satellite DNA could play an important role in gene regulation, thereby affecting adaptive immune response (Pezer et al, 2012).

1.6. The role of satellite DNA transcripts in siRNA-mediated establishment of heterochromatin and gene regulation

Epigenetic research over the past years revealed that satellite DNAs are targets of silencing mechanisms, such as RNA interference (RNAi) and involving epigenetic modifications. Tandem arrangement of satellite DNAs seems to be responsible for triggering RNAi-based silencing mechanisms and formation of heterochromatin, thereby regulating the expression of nearby genes. Expression of nearby genes can as well be influenced by the presence of functional promoters and transcription factor binding sites within satellite sequences, probably through transcriptional interference mechanism. Based on this observations, it is hypothesized that influence of satellite DNAs is epigenetic in its nature and could be modulated by enviornmenal changes (Pezer, 2012).

It has been shown that transcripts derived from tandemly repeated centromeric DNA of fission yeast *Saccharomyces pombe* exist in the form of small 20 - 25 bp long RNAs involved in RNA interference-mediated heterochromatin assembly (Zofall & Grewall, 2006). This mechanism is initiated by small interference RNAs (siRNAs) derived from long double-stranded RNAs (dsRNAs). dsRNAs are produced by bidirectional transcription of repeated DNA and further processed into siRNA by Dicer, a RNase III-like ribonuclease. Through their association with the Argonaute protein, siRNAs are loaded onto RNA-induced transcriptional silencing complex (RITS), which interacts with RNA-directed RNA polymerase complex (RDRC) required for the production of secondary dsRNA and silencing signal amplification. Nascent centromeric RNA transcript binds to RITS through base pairing with siRNA and by direct contact with RNA pol II elongation complex. The association of RITS with chromatin also requires a histone methyltransferase. Methylation of histone H3 at lysine 9 (H3K9) is essential for recruitment of chromodomain protein Swi6, a homologue of mammalian heterochromatin protein 1 (HP1), which represents the initial stem in the formation of heterochromatin and enables further recruitment of heterochromatin assembly factors. Therefore, low satellite DNA expression is surely required to initiate heterochromatic state, but it is still questionable if it is neccesary for its propagation and maintainence.

siRNA expression might as well be involved in chromatin modification, as reported in *D. melanogaster*, where their expression is regulated developmentally and it is most intense in testes and in early embryos. Satellite DNA-derived siRNAs may be involved in post-transcriptional gene regulation through the action of RNA-induced silencing complex (RISC). Several coding mRNAs in human and chick embryos contain alpha-like satellite repeats in their

5'- or 3'- untranslated regions, so there is an indication that their expression could be controled through action of siRNAs derived from alpha satellite repeats (Davidson & Britten, 1979).

It has been hypothesized that the influence of satellite DNA on neighbouring genes is epigenetic in nature and is induced by specific changes in the environment such as long-term heat stress. In the study of modulation of gene expression in *Tribolium castaneum* in the terms of heat stress (Feliciello et al, 2015), the results indicate that satellite DNA-associated siRNAs, transiently activated after heat shock, affect the epigenetic state of dispersed TCAST1 satellite elements inducing their „heterochromatinization". In addition, heterochromatin marks are spread to the regions flanking dispersed satellite DNA elements. The transient increase in H3K9me3/2 levels induces partial repression of genes located in the vicinity and confirms the proposed model of gene regulation mediated by satellite DNA.

In summary, siRNAs originating from satellite repeats seem to have an extensive role in gene expression and heterochromatin formation (Ugarković, 2005; Felicielo, 2015).

1.7. DNA methylation analysis

DNA methylation is a naturally occuring event in both prokaryotic and eukaryotic organisms (Costello & Plass, 2001). It consists of the addition of a methyl group to the fifth carbon position of the cytosine pyrimidine ring via methyltransferase enzyme (Adams, 1995). The role of DNA methylation in prokaryotes is to mark host DNA and make it unsuspeciable to digestion by restriction endonucleases which cleave foreign DNA. In higher eucaryotes, however, DNA methylation acts as a regulating mechanism of gene expression (Costello & Plass, 2001). For instance, DNA methylation has been shown to play a central role in gene imprinting, embryonic development, X-chromosome gene silencing and cell cycle regulation (Adams, 1995). Moreover, abberant DNA methylation is a change that occurs early during oncogenesis and it represents a widespread phenomenon in cancer (Stirzaker, 1997). The majority of DNA methylation occurs in 5'-CpG-3' dinucleotides, even though other methylation patterns do exist. In general, about 80 percent of all CpG dinucleotides in mammalian genome are found to be methylated. The remaining 20 percent happens to be unmethylated and it is primarly located within promoters or in the first exons of genes (Adams, 1995).

A number of methods have been developed to detect/quantify DNA methylation including: high-performance capillary electrophoresis (Fraga et al, 2000) and methylation-sensitive arbitrarly primed PCR (Gonzalgo, 1997). However, the most common method used today in the bisulfite conversion, a technique that involves bisulfite treatment of methylated

DNA which results in conversion of unmethylated cytosines into uracil. Methylated cytosines remain unchanged during the treatment. After conversion, the methylation profile can be determined by PCR amplification followed by DNA sequencing (Frommer, 1992).

1.8. Aims

The aim of this study is to charachterize dispersed elements of alpha satellite DNA in human genome. In this study, I tested existence of polymorphism of alpha satellite elements associated with protein-coding genes in 31 different cell lines and 29 individuals. I analysed alpha satellite elements that are found in the vicinity of genes (close to their 5'- or 3'- end), as well as the elements inserted into introns of genes. I further analysed the methylation profile of regions that contain alpha satellite element in response to heat stress, since it has been proven for satellite DNA to be temperature-sensitive (Feliciello et al, 2013), indicating that it might have role in regulation of gene expression.

2. Materials and methods

2.1. Materials

2.1.1. Solutions, reagents and other materials

- DNA polymerase (DreamTaq Green PCR Master Mix, Thermo Scientific)

- Primers (Table 1) (Macrogen)

- Gel extraction kit (EZ-10 Spin Column DNA Gel Extraction Kit, Bio Basic Inc.)

- Column purification kit (Illustra™ MicroSpin S-200 HR Columns, GE Healthcare)

- Water (AccuGENE, Lonza)

- 1x TAE buffer

    - Concentrated stock solution 50X TAE is diluted to 1X TAE working solution.

        This 1X solution contains in final concentrations 40 mM Tris (Promega), 20 mM acetic acid (Kemika), and 1 mM EDTA, pH 8.0 (Promega)

- Agarose (SeaKem®LE Agarose, Lonza)

- Bisulfite conversion kit (EpiTect® LyseAll Lysis Kit, Qiagen)

- DNA ladder (Quick-Load® 2-Log DNA Ladder 0.1-10 kb (6X), New England BioLabs; FastRuler Middle Range DNA Ladder, Thermo Scientific)

- Loading dye (Gel Loading Dye Purple (6X), New England Biolabs® Inc.)

- Restriction enzymes (*Alu*I, *Hinf*I, *Eco*RI; New England Biolabs)

- Qubit reagents (Qubit® dsDNA BR Reagent * 200X; Qubit® dsDNA BR standard #1 (0 ng/uL, in TE buffer); #2 (100 ng/uL); Life Technologies)

- Ethidium bromide (Ethidium Bromide Solution, Molecular Grade, Promega)

- Basic Local Alignment Search Tool (Nucleotide BLAST, NCBI; https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

2.1.2. Primers for polymorphism analysis

Analysis of polymorphism of dispersed alpha satellite elements was performed by PCR using primers pairs that enabled specific amplification of satellite element located at particular site. Primers were designed using Primer3Plus software. Specificity of primers was assessed using PrimerBlast by blasting chosen primers against NCBI Genome (chromosomes from all organisms) database limiting organism field to *Homo sapiens*. Primer pairs sequences with their marking, temperature of melting (Tm) and expected size of amplicon are listed in Table 3.

All primers were designed according to positions of dispersed alpha satellite elements predicted in NCBI database. For primer design, Primer3 Plus online software was used. Designed primers were subsequently checked for specificity by NCBI's Primer Blast Tool. All primers were ordered from Macrogen, South Korea.

In order to check the presence of a certain satellite element, primers were constructed for each position in the vicinity of the genes or within the introns.

Table 1. A list of primer pairs used in polymorphism analysis. All primers are designed in Primer3 Plus softwer and their specificity is assesed by NCBI Primer Blast online tool.

Primer designation	Element	Primer sequence	Tm / ˚C	Product size / bp
p7039TGFA&ncRNA	2	F: CCACAAATGGGCCCCTGTTA	60,5	547
		R: GGGAACATCACACTCTGGGG	62,5	
p30061&54529	3	F: AGGGCTCAGATTTAGGGATTT	57,4	437
		R: AACCCCATAAGGAAGCCAGT	58,4	
p6480&116832	4	F: CCAGACCACGGTAAATGCCA	60,5	537
		R: GACGGATGGCAGAACAAACG	60,5	
p222584FAM83B	13	F: TAGGAAGTGGGAGAGGGCAA	60,5	593
		R: GACCTGCCCGAAGCTGTTAA	60,5	
p51105&7038	16	F: GGAGTCAGGCAGGTGAGTTC	62,5	552
		R: CATCACCCCATTGGCTGTCT	60,5	
p22982DIP2C&UnPr	17	F: ATGGGGAGGTGATGGGGTTA	60,5	520
		R: CCCTCACCAGAATTGCCCTT	60,5	
pORfamily	20	F: TGCTGCCTACAAAAGACCCA	58,4	577
		R: TCTGGGCTCACTGCAATCTC	60,5	
p28227&miscRNA	30	F: CCTGAGTGGGGAGAAAGCTG	62,5	537
		R: CTGTAGGCCCAGCTACAAGG	62,5	
p2.1 SLC30A6	1	F: GCCTCCTGAGTTCAAGCAAC	64,3	341
		R: GCATGGTGCCTCATTCCTAT	64,2	
p4.1 trimer	5	F: ACAGCAGTGACTCTGAGCCA	64,5	690
		R: AATAGCCTGTTCTTCTGAAGC	59,6	
p4.1_985		F: GTCTGTCTCTGTCTCGGCTG	62,5	985
		R: AAAGATCAGGTCGCGTGTCC	60,5	
p6.2PRIM2	14	F: GTTCCTGTAAGGGCTCAACG	63,9	222
		R: AGGCTGCAGTAAGCCATGAT	64,1	
p7.1 dimer	15	F: AGTAGTGACTCGTAGGACAC	53,8	405
		R: CTTAGCCCCGTCAAGTTGAC	63,9	
p10.2 STAM	18	F: TCCCAGTCCATCGAAACCTA	64,8	248
		R: GAAGCTTCATCACCCTCCAA	64,5	
p10.4 PLA2G12B	19	F: CTTGGGCATTTCTGGTCTGT	64,4	490
		R: ATTAGGGCGTTTCACACAGG	64,2	
P10.4_835		F: ACCCAGTCTCTGTGTCCCAT	60,5	835
		R: GGGCTCCTTTTCAGGGTGAT	60,5	
p11.5 DLG2	21	F: GCACCACACTGGTCTTCC	62,5	355
		R: CTGTATTTCAGCATGAGTGACAG	61,8	
p19.2 ZNF675	29	F: ACAGGTGTGAGCCACTATTC	55,1	325
		R: TAGTAGACACCGGGTTTCACC	59,0	
pNR3C1	10	F: ACCATGTTAGCCAGGATGATC	63,1	361
		R: ACCTCAAGGAAGAACTGGATG	62,5	
pMYO1E	25	F: CGACATGGGTCCAGTCTGAT	61,0	285
		R: AGGAATCTGGATATGTCTTCCA	57,6	
pVav	28	F: AGACTCCATCCCCCTCAAAA	60,8	301
		R: TCAAGGTGTCAGTAGGGTTGG	60,0	
pMAP7D2	31	F: CGCTAATGCTGAAGACATGC	59,6	292
		R: GAAGTGGAATGGGATCTGAAA	59,0	

2.1.3. Primers for pyrosequencing

In order to amplify DNA fragments after bisulfite conversion, primers for PCR and pyrosequencing must be complementary to converted DNA (Li & Dahiya, 2002). Therefore, primers were designed using MethPrimer software. Since minus strand of PCR amplicon served as a template for pyrosequencing reaction, reverse primers were biotynilated on their 5'- end Royo, Hidalgo & Ruiz, 2007). A list of PCR and pyrosequencing primers is shown in Table 2.

Table 2. A list of primers for PCR and pyrosequencing of amplified fragments. All primers are designed complementary to bisulfite converted DNA in MethPrimer online tool.

Name	Sequence (5' -> 3')	Tm / ˚C	
ZNF675Pyro	TAATGTAGTTTTGAAATTATAAAGG	57.89	**Sequencing**
NR3C1Pyro1	ATTATGTTAGTTAGGATGA	53.45	
NR3C1Pyro2	GAATTTGTAAGTGGATATTTG	56.63	
ZNF675Fw	ATAGGTGTGAGTTATTATT	53.77	**PCR**
ZNF675Rev	5' Btn-CTTCTTTCTAATTTTTATCTAAC	55.15	
NR3C1Fw1	ATTATGTTAGTTAGGATGATT	55.55	
NR3C1Fw2	ATAGAATTTGTAAGTGGATATT	56.57	
NR3C1Rev1	Btn-CCAAATATCCACTTACAAATTCT	60.85	
NR3C1Rev2	Btn-ACCTCAAAAAAAAACTAAATAAATTTT	60.29	

2.1.4. Cell lines

To analize polymorphism in cell lines, a DNA isolated from 27 different cell lines was used (Table 3.).

Table 3. DNA isolated from 27 various cell lines. Concentration of DNA for each sample was measured by Qubit fluorometer (Broad Range).

Abreviation	Cell Line	c (ng/uL)
MJ90	Human fibroblast	55,3
MCF	Human breast adenocarcinoma	72,3
HeLa	Human cervical cancer	150
Hep2	HeLa derivative	112
SW620	Human colon cancer	76,3
DLD1	Human colorectal adenocarcinoma, Dukes type C	44
HCT116	Human colorectal carcinoma	130
ZR75	Human ductal carcinoma	61,2
HepG2	Human hepatocelular carcinoma	106
A431	Human epidermoid carcinoma	144
SW48	Human colorectal carcinoma, Dukes type C, grade IV	95,4
697	Human B cell precursor leukemia	9,32
Cal27	Human squamous cell carcinoma	2,24
K1 IMRO	Human thyroid carcinoma	22,9
293T	Human embrionic kidney	18,7
PBL	Human peripheral blood lymphocytes	15,3
HEp2	Human laryngeal carcinoma, epithelyal cells type 2	32,5
MES-OV	Human ovarian cancer	48,3
LNCaP	Androgen-sensitive human prostate adenocarcinoma	22,5
OV-90	Human ovarian adenocarcinoma	16,7
PC-3	Human prostate adenocarcinoma	55,0
Mewo	Human melanoma	17,9
SH-SY5Y	Human neuroblastoma	25,8
MDA MB231	Mammary gland / breast adenocarcinoma	73,8
H4	Neuroglioma	11,2
Jurkat	Acute T cell leukemia	84,0
A1235 At	Human glioblastoma cell line transfected with TGFβ	79,1

2.1.5. Peripheral blood leukocytes (PBL)

To test the presence of dispersed elements and occurence of polymorpshims in the population, DNA from human peripheral blood leukocytes (PBL) was isolated (Table 4). The blood samples were given voluntairily by 26 female patients for the needs of scientific research at Ruđer Bošković Institute. The isolated DNA used in this experiment was recived with designations which do not reveal identity, age or any other informations about patients.

Table 4. DNA isolated from 26 samples of human peripheral blood leukocytes (PBL). Concentration of DNA for each sample was measured by Qubit® fluorometer (Broad Range).

Sample	c (ng / uL)
M1	13,5
M2	36,4
M3	21,1
R2	55,8
R3100	99,8
R3200	102, 0
R4	437, 0
R5	50,3
R6	42,8
R7	23,5
R19	331, 0
16	23,5
Z20	37,4
R8	62,0
R13	7,9
R14	16,4
R17	24,1
R20	21,5
R21	26,7
R23	21,2
R33	40,4
R43	58,8
13	40,1
24	17,1
Z21	26,3
Z23	26,9

2.1.6. Human fibroblast cell culture

hTERT MJ90 (human neonatal foreskin fibroblasts with activated telomerase) cells were provided by Laboratory for Molecular and Cell Biology, under supervision of Dr. Ivica Rubelj, Assoc. Prof. Cell culture was grown in Dullbecco's Minimum Eagle's Medium (DMEM) with 10 % fetal calf serum (FCS) and gentamycine in T25 flasks, at 37 ˚C and 5 % CO_2, untill it reached confluency of 70 %.

2.2. Methods

2.2.1. PCR analysis

PCR reactions were made using premade 2x DreamTaq Green PCR Master Mix in a final reaction volume of 20 µL. Final concentration of specific primers in reaction was 0,2 µM. 30 – 100 ng of genomic DNA was added to each reaction tube. To yield higher amount of product needed to perform PCR purification or agarose gel purification step, reaction volume was increased to 40 µL or 50 µL, followed by the increase of amount of DNA added. PCR was performed on Biometra® T Personal machine. Programme is described in Table 5. In order to selectively amplify fragments of interest, the annealing temperature was lowered for 0,4 ˚C per each of 10 cycles („touchdown“ PCR). The default annealing time (20 s) was modified according to the lenght of the desired product.

Table 5. PCR programme for polymorphism analysis

Step	Temperature	Duration	Number of cycles
Pre-heating	94 °C	1 min	
Denaturation	94 °C	30 s	10 x
Annealing	60 °C	20 s	
Elongation	70 °C	20 s	
Denaturation	94 °C	30 s	30 x
Annealing	55 °C	10 s	
Elongation	70 °C	20 s	
Final elongation	70 °C	1 min	
	4 °C	infinite	

PCR products were subsequently checked by agarose gel electrophoresis. Most PCR products were visualized on 1,0 % agarose gel stained by ethidium bromide. 1,0 % agarose gel was prepared by dissolving 1,2 g of LE agarose (Roche) in 100 mL Tris-acetate-EDTA (TAE) buffer containing ethidium bromide in final concentration 1 µg/mL. A higher percantage of agarose (1,2 to 1,8 %) was used to separate particular fragments better. In contrast, if gel purification step needed to be performed, it was necessary to use preparative gels containing the lower percentage of agarose (for example, 0,8 %). All 20 µL of reaction were loaded on the

gel, except when column purification of PCR products needed to be performed. To assess the size of DNA molecules after electrophoresis, Quick-Load® 2-Log DNA Ladder (0.1-10 kb) (6X), New England BioLabs and FastRuler Middle Range DNA Ladder; Thermo Scientific were used. Electrophoresis was performed at voltage in the range from 70 V to 120 V for 45 - 60 min. Gel was visualized under transiluminator using wavelengths in the range from 500 nm to 700 nm, depending on the amount of visible products. Photography of gel was taken and transferred to computer using Gene Snap (Syngen) program, v. 7. 05. All photograpies were furtherly processed and edited using Photoscape photo editing tool.

2.2.2. Enzymatic digestion

In order to remove the unwanted fragment or to check the presence of alpha satellite inside the fragment, some genomic DNA samples or PCR products were treated with restriction enzymes.

Enzymes *Alu*I and *Hinf*I were used to digest element 29 (associated with gene *ZNF675*) in samples R4 and *Alu*I + *Eco*RI were used to digest R4's genomic DNA before performing a PCR reaction with primers specific for element 29.

The digestion step for all digestion reactions is performed as following: 1 μL of enzyme (5 U/μL) was added per 10 μL of reaction, then put at 37 ˚C for 1 h. Since the enzyme is compatible with the Dream Taq buffer used in previous PCR step, enzyme was added directly to PCR product. It is important to note that enzyme was added in excess (1 μL can digest up to 1 μg of DNA).

2.2.3. Column purification of PCR product

To remove the DNA polymerase, nucleotides and other reagents left in the solution after PCR, PCR products were transfered to Ilustra™ MicroSpin TM S-200 HR Columns, GE Healthcare. Each column contains resins that need to be mixed thoroughly by vortexing before use. To prepare the column for binding, a pre-requisite step is centrifuge at low speed (735 x g) for 1 min to get rid of the liquid portion. Before centrifuging, it is important to loose the cap (approximatley one quater turn) and to twist off the bottom closure. After the preparation, the flow-through was discarded and each column was transfered to clean 1.5 mL Eppendorf tubes.

25 – 100 μL of PCR product was transfered carefully by pipetting without disturbig the resin's surface and centrifuged at 735 x g for 2 min. Eluted DNA was stored at -20 ˚C.

2.2.4. Gel extraction

To extract DNA from the agarose gel, a band representing a DNA fragment of interest was cut out of gel under UV using sharp scalpell and transfered to clean 1,5 mL Eppendorf tube. 400 μL of Binding Buffer II per 100 mg of gel was added onto gel band. If the gel contains higher percentage of agarose (1,5 – 2 %), it is neccesarry to add 500 μL per 125 mg of gel. To dissolve the gel, the tube was incubated at 50 – 60 ˚C for 10 min with occasional mixing by shaking or vortexing. The mixture was transfered to EZ-10 columns and left stand for 2 min and then centrifuged at 10 000 rpm for 2 min. After performing two washing steps using 750 μL of Wash solution with a subsequent centrifugation at 10 000 rpm, each column was transfered to a clean 1,5 μL Eppendorf tube. 30 – 50 μL of Elution Buffer buffer was added and left stand 2 min at RT and centrifuged at 10 000 rpm for 2 min. Eluted purified DNA was stored at -20 ˚C.

2.2.5. DNA sequencing by Sanger method

In order to acess the sequence of short fragments, a commonly used method is Sanger sequencing by synthesis. Each time fluorescent dedioxyribonucleotydes complementary to the DNA template are being incorporated during DNA polymerization, the synthesis process is being paused, thus generating fragments of different lenghts. Fragments are then being separated on capilary electrophoresis, which can separate fragments differing in 1 bp, and generated as peaks. Final results of sequencing are exported as chromatograms.

DNA sequencing was conducted at the sequencing facility at Ruđer Bošković Institute. The sequencing reaction proceeds in two directions (forward and reverse) so two reaction tubes are needed for each sample. To prepare samples for sequencing, 50 – 500 ng of DNA needs to be mixed with 1 μL of sequencing primer (c = 3,2 μmol/μL) in maximal reaction volume of 13 μL. Therefore, all samples were prepared for sequencing by mixing 1 μL of primer and 10 μL of DNA.

2.2.6. Heat-shock treatment

For the heat-shock treatement and control experiments, two incubators set each on different temperature (37 ˚C, 42 ˚C) were used. Four T25 flasks containing hTERT MJ90 cells at 70 % confluency were treated as described in Table 6.

Table 6. Heat shock treatment. Four T25 flasks containing hTERT MJ90 cells (70% confluency) were divided in 3 groups, where 1st group was control group, the second one was treated group and the third one recovered after heat-shock treatment.

Flask	Temperature	Duration	Description
1st	37 ˚C	2 h	**Control treatment**
2nd	42 ˚C	2 h	**Heat-shock treatment**
3rd	42 ˚C	2 h	**Heat-shock treatment (duplicate)**
4th	42 ˚C	2 h	**Heat-shock treatment**
	37 ˚C	1 h	**Recovery**

Treatments were performed simoultaniously. Each treatment was immediatley followed by DNA isolation step.

2.2.7. Isolation of total DNA from human fibroblas cell culture

Total human genomic DNA was isolated from treated human fibroblast cells using GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich).

The growth medium was removed by aspiring and cells were lysed by adding 200 µL of Lysis Solution directly in the flask. Attached cells were then scrapped into lysis solution and transfered to clean Eppendorf 1,5 mL tubes. 20 µL of Proteinase K was added to each tube, then vortexed and put at 70 ˚C for 10 min. Afterwards, 20 µL of RNAseA was added and left for 2

min at room temperature. For column preparation, 500 μL of Column Preparation Solution was added to each preasambled GenElute Miniprep Binding Column, and centrifuged at 12 000 x g for 1 min. Lysate was prepared for binding by adding 200 uL of 95% ethanol and vortexing for 5 to 10 s. Lysate was transfered to the binding column and centrifuged at 6500 x g for 1 min. Two washing steps were performed using 500 uL of Wash Solution Concentrate (with added ethanol) per each wash step and centrifuging: first at 6500 x g for 1 min, and then at maximum speed (16 000 x g) for 3 min. After each centrifuge step, the liquid flow-through was discarded. To make sure the column is ethanol free, the binding column with an empty collection tube was centrifuged at maximum speed (16 000 x g) for 1 min. The ethanol leftovers were discarded carefully and the binding column was transfered to 1,5 mL Eppendorf tube. To elute DNA from column, 200 μL of Elution Buffer was added and left for 5 min at RT to increase the elution efficiency. DNA was eluted by performing a centrifuge step at low speed (6500 x g) for 1 min. Eluted DNA was stored at 4 ˚C.

2.2.8. Measurement of DNA concentration

For all DNA samples, DNA concentration was measured using Qubit dsDNA Broad Range (BR) Assay kit (Invitrogen). The measurement is performed on Qubit™ Fluorometer using 1 μL of DNA sample mixed with 199 μL of solution made by mixing 199 μL of BR Buffer and 1 μL of fluorescent reagent Qubit® dsDNA BR Reagent * 200X. Before every measurment run, the fluorometer was calibrated using two standard solutions (10 μL of Qubit® dsDNA BR standard #1 (0 ng/uL, in TE buffer); #2 (100 ng/uL) mixed with 190 μL of solution made by mixing 199 μL of BR Buffer and 1 μL of fluorescent reagent).

2.2.9. Bisulfite conversion

The method was first introduced by Frommer et al (1992) and it is based on the difference between the products of amination reactions of cytosine and 5-methylcytosine. Particulary, after the treatment with sodium bisulfite, cytosines in single-stranded DNA will be converted into uracil residues, while 5mC will remain immune to this conversion yielding cytosines. In a subsequent PCR amplification and sequencing, uracil residues will be recognized and amplified as thymines, which will distinguish methylated from unmethylated cytosines. A subsequent PCR process is necessary to determine the methylation status at the loci of interest

by using specific methylation primers after the bisulfite treatment. DNA methylation can be interpreted by subsequent sequencing analysis.

Bisulfite treatment was performed using EZ DNA Methylation-Gold Kit (Zymo Research) following the protocol provided by manufacturer. Before starting the conversion, the CT Conversion Reagent was prepared by adding 900 µL of water 50 µL of M-Dissolving Buffer and 300 µL of M-Dilution Buffer to one tube of CT Conversion Reagent and mixed for 10 min. 130 µL of prepared reagent was added to 20 µL of DNA sample and mixed. 100 ng of DNA was used for each reaction. Bisulfite conversion requires elevated temperature so the conversion reaction was carried out by designed temperature programme (Table 7).

Table 7. Temperature programme for bisulfite conversion reaction

Step	Temperature	Duration
1.	98 ˚C	10 min
2.	6 ˚C	2,5 h
3.	4 ˚C	infinite

Six hundred µl of M-binding Buffer was added to a Zymo-Spin IC Column and then the sample was transfered. The solution was mixed by inverting several times. Centrifuge step was performed at full speed (over 10 000 x g) for 30 sec and the flow-through was discarded. 100 µL of M-Wash Buffer was added to the column and spined for 30 sec. 200 µL of M-Desulphonation Buffer was added to the column and left sitting for 15-20 min. Spin was performed at full speed for 30 sec. The column was washed twice by adding 200 µL of M-Wash Buffer and spining for 30 sec. The column was transfered to a new clean 1,5 mL Eppendorf tube. In order to elute DNA, 10 µL of M-Elution Buffer was added directly to the column matrix and spinned briefly.

2.2.10. Pyrosequencing

Pyrosequencing is a real-time DNA-sequencing method based on the detection of the pyrophosphate (PPi) released during the DNA polymerization reaction (Nyrén & Lundin, 1985; Hyman, 1988; Ronaghi et al, 1996). Once the PCR products are converted into single-stranded DNA (ssDNA) fragments and one strand is isolated (usually through labeling with biotin) for use as template in the pyrosequencing reaction, a sequencing primer is incorporated into the

reaction and a specific nucleotide is dispensed. When pyrophosphate molecules are released upon incorporation of the nucleotide in the elongating strand, ATP-sulfurylase converts it to ATP, thus providing the energy for luciferase, that catalyses luciferin into oxyluciferin (Royo, Hidalgo & Ruiz, 2007). The average number of emitted photons per template chain in a given step is proportional to the number of deoxynucleotides incorporated per chain at that step (this relation is linear only for a small number of incorporations). The sequence can then be determined by monitoring if incorporations occur and by counting the number of incorporations (by measuring the light intensity) in a given attempt. The amount of light emitted can be measured by an avalanche photodiode, photomultiplier tube, or with a charged-coupled device camera (with or without a microchannel plate) (França, Carrilho & Kist, 2002) (Figure 3).

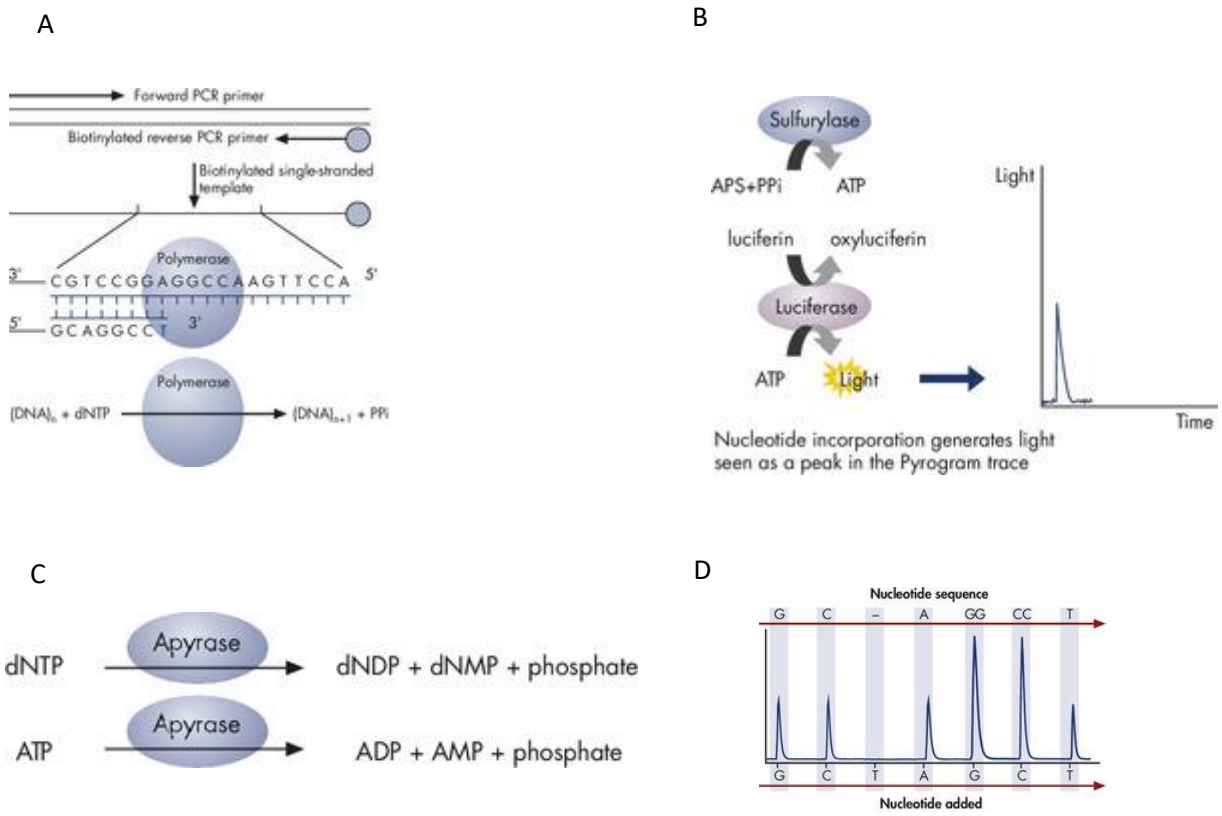


Figure 3. A schematic representation of pyrosequencing mehod. A. PCR generates biotinylated products. B. ATP generated by sulfurylase emits light signal in luciferase-mediated redox reaction C. Apyrase degrades redundant dNTP and ATP D. Results of pyrosequencing are generated as pyrograms. Peak height is proportional to the number of particular nucleotide incorporated (taken and modified from https://www.qiagen.com/dk/resources/technologies/pyrosequencing-resource-center/technology-overview/)

The fragments for pyrosequencing are amplified by PCR using primers designed complementary to bisulfite converted DNA. Fragments of interest contain alpha satellite elements located in introns of *NR3C1* (element 10) and *ZNF675* (element 29) (see chapter Results). All reverse primers were biotynilated on their 5'- end. Since the element 10 was too big for pyrosequencing, it was divided in 2 fragments which have been analysed as two separate pyrosequencing runs.

Pyrosequencing was performed on Pyromark Q24 (Qiagen). This facility was generously provided by laboratory of Dr. Vlatka Zoldoš, Prof., at Department of Molecular Biology, Faculty of Science, University of Zagreb, Croatia. A biotinylated PCR product is bound to streptavidin-coated sepharose beads which were captured with the vacuum tool on the vacuum workstation. After the capture, the beads with bounded products were thoroughly washed and subsequently denatured. Denaturation generates a single stranded DNA which is used as a sequencing template which is being released into the pyrosequencing reaction plate containing the sequencing primer. After primer annealing, the plate was placed into the PyroMark instrument. All reagents (enzymes, nucleotides, and substrate for the pyrosequencing reaction) were pipetted into the dispensing tips or cartridge (depending on the instrument used), according to the volumes provided by the software, and are also placed into the instrument for the pyrosequencing run. The results of pyrosequencing reaction are exported as pyrograms.

2.2.11. Statistical analysis

One-way ANOVA

In statistics, one-way analysis of variance (abbreviated one-way ANOVA) is a technique that compares means of two or more samples (using the F distribution). It can be used only for numerical response data, the "Y", usually one variable, and numerical or (usually) categorical input data, the "X", always one variable, hence "one-way" (Howel, 2002).

General asumptions of one-way ANOVA include normal or approximately normal distribution of independent samples and equal variance of populations. The null hypotesis of one-way ANOVA test is that all population means are equal.

The Kruskal–Wallis test by ranks, Kruskal–Wallis H test (named after William Kruskal and W. Allen Wallis), or one-way ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or

more independent samples of equal or different sample sizes. Since it is a non-parametric method, the Kruskal–Wallis test does not assume a normal distribution of the residuals, unlike the analogous one-way analysis of variance (Siegel & Castellan, 1988).

ANOVA partitions the variability among all the values into one component that is due to variability among group means (due to the treatment) and another component that is due to variability within the groups (also called residual variation). Variability within groups (within the columns) is quantified as the sum of squares of the differences between each value and its group mean (residual sum-of-squares). Variation among groups (due to treatment) is quantified as the sum of the squares of the differences between the group means and the grand mean (the mean of all values in all groups). (Bewick, Cheek & Balls, 2004). Adjusted for the size of each group, this becomes the treatment sum-of-squares. Each sum-of-squares is associated with a certain number of degrees of freedom (df) and the mean square (MS) computed by dividing the sum-of-squares by the appropriate number of degrees of freedom.

To statisticaly test the significance of the observed differences in level of methylaton for different treatments, one-way ANOVA with non-Gauss distribution was performed in GraphPad Prism 7.

3. Results

3.1. Localization of dispersed alpha satellite elements in euchromatin

To check the dispersion of alpha satellite repeats within euchromatin, we searched the human reference genome GRCh38.p7 with the alignment program BLASTN version 2.3.1+. using as a query human alphoid 171 bp repeat consensus sequence. Only hits with at least 50% of repeat sequence length and > 70% identity to the query sequence were considered for further analysis. Except chromosomes with exclusive clustered organization of alpha satellite repeats, 13 chromosomes have in addition to clusters, single repeats or short arrays up to a trimer of alpha satellite DNA dispersed at intergenic regions or within introns. There are in total 31 dispersed alpha satellite DNA elements, of which 10 are within introns while the remaining 21 are within intergenic regions. Such an organization of dispersed alpha-like repeats within introns of genes, as well as within intergenic regions suggests their possible influence on the expression of adjacent genes. In order to test the potential gene-regulatory role of dispersed alpha satellite elements we first analysed their insertion polymorphism among different human cell lines and human individuals. Analysis of insertion polymorphism was performed for all 10 alpha satellite elements located within introns as well as for 10 elements in intergenic regions which are located within 50 kb relative to the associated gene (Table 8).

Table 8. Localization of dispersed alpha satellite elements in euchromatin. Analysis of insertion polymorphism was parformed for 10 elements located within introns and for 10 elements in intergenic regions (bold).

Chr No	alpha repeat No	Genome position	Similarity to alpha satellite consensus/size	associated gene
2	**1**	**32213935-32214051**	**82%, 0,7 monomer**	**solute carrier family 30 member 6, (SLC30A6 ID: 55676), intron**
	2	**70431412-70431596**	**88%, 0,9 monomer**	130438 bp at 5' side: protein FAM136A isoform X1, ID: 84908
				19263 bp at 3' side: TGFA transforming growth factor alpha isoform 2 preproprotein, (TGFA ID: 7039)
	3	**189619699-189619943**	**77%, 1,4 monomers**	**38888 bp at 5' side: solute carrier family 40 member 1 isoform X1, (SLC40A1 ID: 30061)**
				46190 bp at 3' side: asparagine synthetase domain-containing protein 1 isoform X1, (ASNSD1 ID: 54529)
3	**4**	**187105323-187105556**	**70%, 1,4 monomers**	**29520 bp at 5' side: beta-galactoside alpha-2,6-sialyltransferase 1 isoform b, (ST6GAL1 ID:6480)**
				15589 bp at 3' side: 60S ribosomal protein L39-like, (RPL392 ID: 116832)
4	**5**	**8342053-8342633**	**75%, 3,4 monomers**	**34942 bp at 5' side: HtrA serine peptidase 3, (HTRA3 ID: 94031)**
				24328 bp at 3' side: peroxisomal acyl-coenzyme A oxidase 3 isoform X4, (ACOX3 ID: 8310)
	6	77489446-77489557	82%, 0,7 monomer	323522 bp at 5' side: cyclin-G2 isoform X1, ID:901
				116309 bp at 3' side: C-X-C motif chemokine 13 precursor ID: 10563
	7	111490831-111490919	78%, 0,52 monomer	857833 bp at 5' side: PITX2 paired like homeodomain 2:, ID: 5308
				335840 bp at 3' side: ncRNA, ID: 105377368
	8	135576694-135576783	89%, 0,52 monomer	1375501 bp at 5' side: polyadenylate-binding protein 4-like, ID: 132430
				1944246 bp at 3' side: protocadherin-18 isoform 2 precursor, ID: 54510
5	9	100196524-100196622	89%,, 0,6 monomer	671163 bp at 5' side: putative POM121-like protein 1-like isoform X2, ID:100652833
				192353 bp at 3' side: putative POM121-like protein 1-like, ID:441098
	10	**143309356-143309541**	**88%, 1,1 monomers**	**nuclear receptor subfamily 3 group C member 1(*NR3C1*, ID: 2908), intron**
	11	161040048-161040173	73%, 0,73 monomer	354828 bp at 5' side: probable phospholipid-transporting ATPase VB isoform X3; ID: 23120
				253908 bp at 3' side: gamma-aminobutyric acid receptor subunit beta-2 isoform 2, ID:2561
	12	161163077-161163193	70%, 0,68 monomer	477857 bp at 5' side: probable phospholipid-transporting ATPase VB isoform X3, ID: 23120
				130888 bp at 3' side: gamma-aminobutyric acid receptor subunit beta-2 isoform 2, ID:2561
6	**13**	**54991256-54991369**	**74%, 0,7 monomer**	**49249 bp at 5' side: protein FAM83B isoform X1; (FAM83B ID: 222584)**
				183219 bp at 3' side: orexin receptor type 2; ID: 3062
	14	**57377966-57378055**	**70%, 0,5 monomer**	**DNA primase subunit 2 (PRIM2 ID: 5558) intron**
7	**15**	**1519232-1519571**	**86%, 2 monomers**	**14850 bp at 5' side: integrator complex subunit 1 (INTS ID: 26173)**
				19622 bp at 3' side: transcription factor MafK, (MAFK ID: 7975)
8	**16**	**132864579-132864640**	**72%, 0,5 monomer**	**15772 bp at 5' side: PHD finger protein 20-like protein 1 isoform X10; (PHF20L1 ID: 51105)**

				2361 bp at 3' side: thyroglobulin precursor, (TG ID: 7038)
10	**17**	**707868-707956**	**94%, 0,5 monomer**	**18158 bp at 5' side: DIP2C disco interacting protein 2 homolog C; (DIP2C ID: 22982)**
				104970 bp at 3' side: la-related protein 4B isoform X2; ID: 23185
	18	**17653288-17653378**	**76%, 0,5 monomer**	**signal transducing adapter molecule 1; (STAM ID: 8027), intron**
	19	**72939337-72939702**	**74%, 2 monomer**	**phospholipase A2 group XIIB; (PLA2G12B, ID: 84647), intron**
11	**20**	**6810439-6810612**	**86%, 1 monomer**	**14731 bp at 5' side: olfactory receptor 6A2, (OR6A2 ID:338755)**
				35071 bp at 3' side: olfactory receptor 10A5, (OR10A5 ID: 338755)
	21	**85431074-85431184**	**85%, 0,7 monomer**	**discs large homolog 2, (DLG2 ID: 1740), intron**
13	22	23423970-23424059	71%, 0,5 monomer	12731 bp at 5' side: sacsin isoform X2; ID: 26278
				166125 bp at 3' side: tumor necrosis factor receptor superfamily member 19 isof; ID: 55504
	23	23424792-23424872	71%, 0,5 monomer	13553 bp at 5' side: sacsin isoform X2, ID: 26278
				165312 bp at 3' side: tumor necrosis factor receptor superfamily member 19 isof., ID: 55504
15	24	24754024-24754329	70%, 1,8 monomers	74686 bp at 5' side: nuclear pore-associated protein 1, ID: 23742
				200720 bp at 3' side: SNRPN upstream reading frame protein,ID: 6638
	25	**59219764-59219973**	**88%, 1,2 monomers**	**myosin IE (MYO1E, ID: 4643), intron**
	26	93907560-93907673	79%, 0,7 monomer	700821 bp at 5' side: putative uncharacterized protein UNQ9370/PRO34162; ID: 105370980
				390593 bp at 3' side: MCTP2 multiple C2 domains, transmembrane 2, ID: 55784
	27	93907750-93907849	77%, 0,6 monomer	701011 bp at 5' side: putative uncharacterized protein UNQ9370/PRO34162, ID: 105370980
				390417 bp at 3' side: multiple C2 and transmembrane domain-containing protein 2, ID: 55784
19	**28**	**6819160-6819369**	**70%, 1,2 monomers**	**vav guanine nucleotide exchange factor 1, (VAV, ID: 7409), intron**
	29	**23661142-23661242**	**86%, 0,6 monomer**	**zinc finger protein 675, (*ZNF675*, ID: 171392), intron**
X	**30**	**402751-402878**	**70%, 1 monomer**	**16060 bp at 5' side: PPP2R3B protein phosphatase 2 regulatory subunit B''beta, (PPP2R3B ID: 28227)**
				228020 bp at 3' side: short stature homeobox protein isoform SHOXb, ID: 6473
	31	**20101282-20101513**	**71%, 1,4 monomers**	**MAP7 domain containing 2 (MAP7D2, ID: 256714), intron**

3.2. Analysis of insertion polymorphism in cell lines

The analysis of insertion polymorphism of alpha satelite elements was carried out in 27 various cell lines. Genomic DNA isolated from cell lines was used as a template for amplification of desired fragments using specific primers which encompass 5' and 3' region of each alpha satellite element. The primers are listed in Table 1.

After the first PCR run, most of the samples showed amplification of a single fragment whose size corresponds to expected size of a fragment that contains alpha satellite element, as predicted in database. A few samples showed unspecific amplification present as faint bands on agarose gel, along with the specific amplification of predicted fragments (element 10). Moreover, some samples showed no amplification at particular loci. To leave out the possiblity of experimental mistake, reactions of interest were repeated. In the case of element 20, reactions failed to gain any products in most of the cell lines even after repeated reaction. None of the amplified fragments indicated insertion polymorphism in alpha satellite element (Table 9).

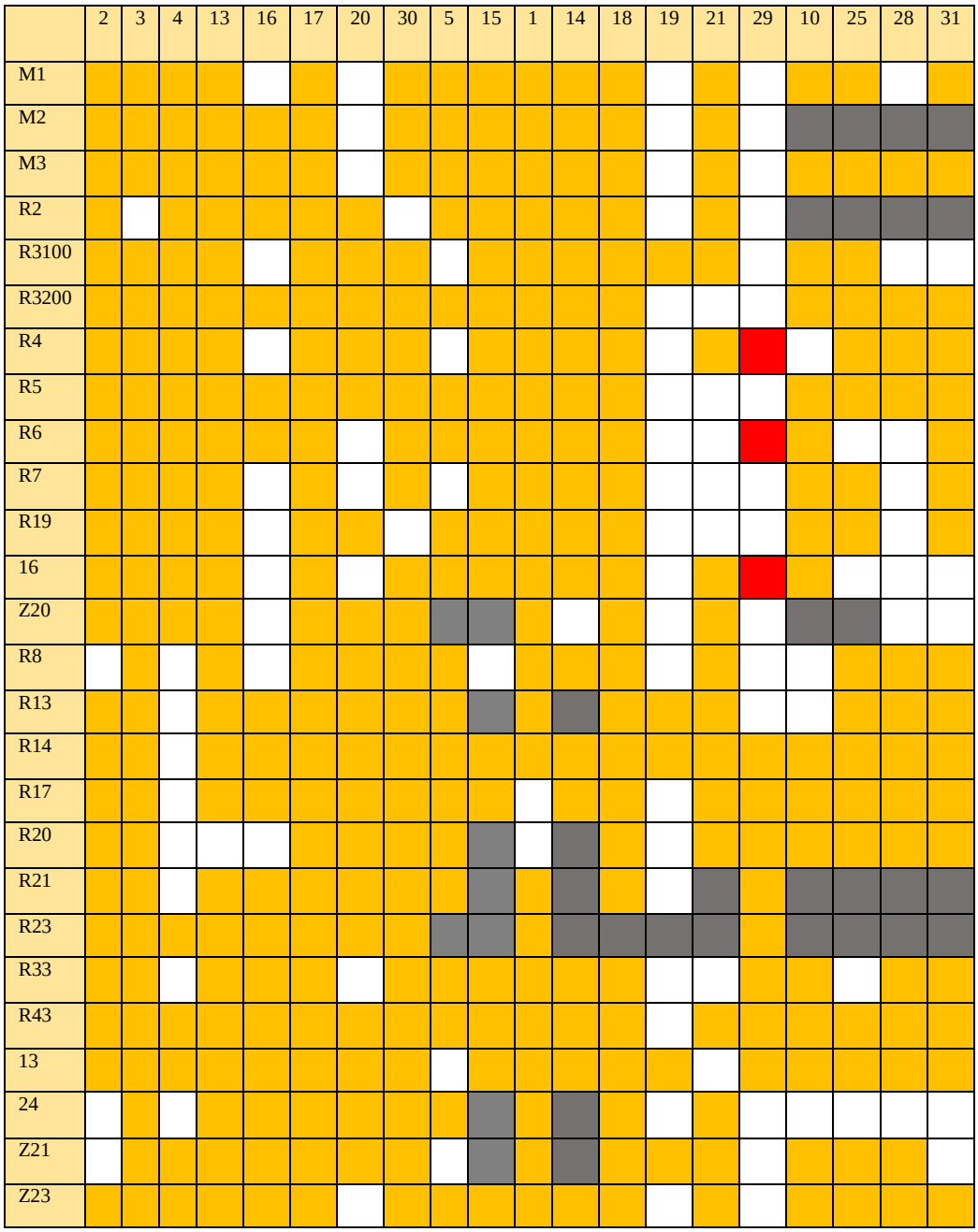
Table 9. Results of insertion polymorphism of alpha satellite repeats in 27 various cell lines. Alpha satellite repeats are indicated by numbers according to Table 8. Each box represent one PCR reaction carried out with particular primer pair (column) and particular DNA sample (row) isolated from cell line. Green boxes represent succesfully amplified fragmens whose size corresponds to the expected size of fragments that contain alpha satellite element. Blank white boxes indicate that no amplification occured. Grey boxes are reactions that had not been carried out.

	2	3	4	13	16	17	20	30	5	15	19	1	14	18	21	29	10
MJ90																	
MCF																	
HeLa																	
Hep2																	
SW620																	
DLD1																	
HCT116																	
ZR75																	
HepG2																	
A431																	
SW48																	
697																	
Cal27																	
K1 IMRO																	
293T																	
PBL																	
HEp2																	unspec
MES-OV																	unspec
LNCaP																	unspec
OV-90																	unspec
PC-3																	unspec
Mewo																	unspec
SH-SY5Y																	unspec
MD Mb231																	unspec
H4																	unspec
Jurkat																	unspec
A1235 At																	unspec

3.3. Analysis of insertion polymorphism in human peripheral blood leukocytes

The analysis of insertion polymorphism was performed on human genomic DNA isolated from peripheral blood leukocytes of 26 individuals. Just as previous analysis in cell lines, most of the fragments in intergenic regions were successfully amplified and their size corresponds to database information, which indicates the presence of alpha satellite element. Still, some samples show the complete absence of amplification at certain loci. The analysis of alpha satellite elements in introns revealed the amplification of fragments which contain alpha satelite elements, as well as the complete absence of amplification at some loci. However, three samples (R4, R6 and R14) showed clear amplification of 2 fragments within intron of *ZNF675* gene (Figure 4.) which correspond to alpha repeat number 29. Size of one fragment corresponds to database predicted size and the size of the other, shorter fragment differs by approximately 150 bp and it might resemble a deletion of alpha satellite element. Results of amplification reactions were summarized in Table 10.

Table 10. Results of insertion polymorphism analysis of alpha satellite repeats in human peripheral blood leukocytes. Blood samples were collected from 26 individuals and genomic DNA was isolated. Alpha satellite repeats are indicated by numbers according to Table 8. Each box represent one PCR reaction carried out with particular primer pair (column) and particular DNA sample (row) isolated from cell line. Orange boxes represent succesfully amplified fragmens whose size corresponds to the expected size of fragments that contain alpha satellite element. Blank white boxes indicate that no amplification occured. Red boxes represent only 3 samples that exhibit different pattern of amplification and contain 2 fragments that differ by the approximate size of element 29 (located in intron of gene *ZNF675*). Grey boxes are reactions that had not been carried out due to the low amount of DNA.



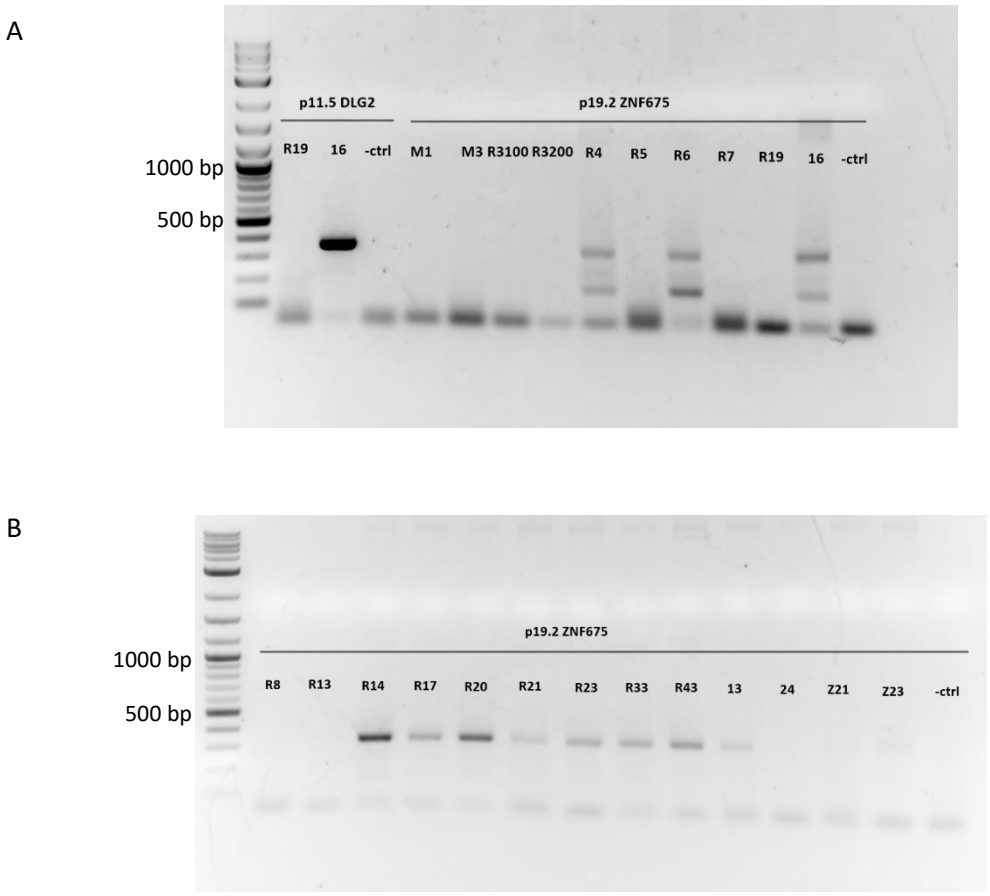


Figure 4. Analysis of polymorphism in element 29 located in intron of gene *ZNF675*. Most of the samples (B) exhibit clear pattern of amplification of one fragment or show no amplification. Samples R4, R6 and R14 show amplification of 2 fragments (A). The shorter fragment is assumed to lack alpha satellite element.

In order to examine the absence of alpha satellite element and determine a polymorphism, the reaction was repeated for samples R4, R6 and 16. The repeated PCR reaction and the subsequential electrophoresis run failed to show two clear separate bands like in the first reaction. In order to isolate the fragment that possibly lacks the alpha satellite element, DNA purified directly from PCR product of R4 sample was treated with restriction enzymes *Alu*I and *Hinf*I that cut inside the alpha satellite element (Figures 5, 6). After the digestion, a portion of digestion product was used as a template for PCR with primers specific for element 29. The idea of this strategy was to remove the amplicon that contains element 29 and to selectively re-amplify the shorter amplicon which has a possible deletion of alpha satellite element and thus does not contain *Alu*I and *Hinf*I recognition sites. The alternative strategy was carried out by first treating genomic DNA from R4 sample with *Alu*I and *Eco*RI restriction endonucleases and then using the digestion product as a template for a PCR reaction with primers specific for element 29.

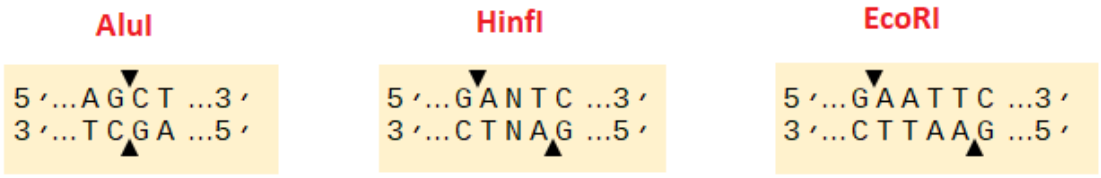


Figure 5. Palindromic recognition sequences of *Alu*I, *Hinf*I and *Eco*RI restriction endonucleases.

Figure 6.. Enzyme digestion reaction of a PCR product containing alpha satellite element. Restriction endonucleases *Alu*I and *Hinf*I cut inside the alpha satellite.

Gel electrophoresis revealed that sucessfull amplification of a desired shorter fragment was achived by performing the first strategy, where PCR product digested by *Alu*I and *Hinf*I was used as a template for re-amplification (Figure 7). The band of interest was extracted from the gel. The extracted DNA was sequenced and aligned with the referent sequence (GenBank, ID: 23361, assembly GRCh38.p7). The alignment revealed that the shortened fragment originated from 2 deletion events. A significant rearrangement of analysed region and absence of portion of alpha satellite element is clearly visible. In addition, the shortened fragment was compared with R4 amplicon that contains alpha satellite element 29 and is refered to as „longer fragment“. The longer fragment was isolated from the gel and directly from PCR product that contained re-amplified longer fragment. Like the shortened fragment, the longer fragment was sequenced by Sanger method.

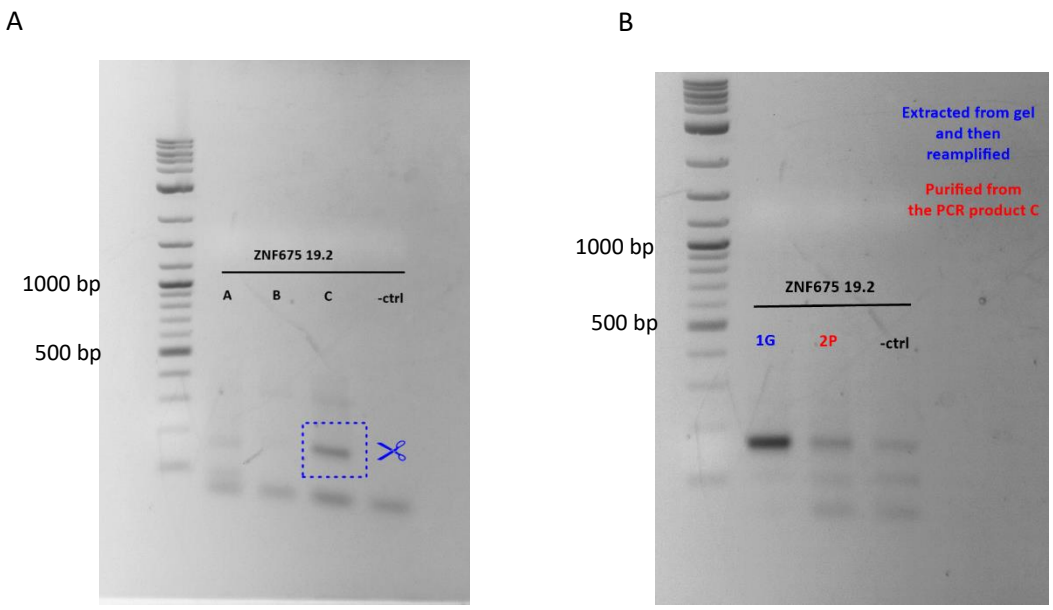


Figure 7. Gel extraction of DNA fragment of interest. A. Sample A, B and C represent three diferent strategies of selective amplification of a DNA fragment that does not contain alpha satellite element. Sample A is R4 amplification product treated with *Alu*I and *Hinf*I. Sample B is R4 genomic DNA treated with *Alu*I and *Eco*RI. Sample C is the product of re-amplification of sample A with primers specific for element 29. Fragment of interest (C) was cut from gel and the DNA was purified with gel extraction kit (Bio Basic Inc). B. The isolated fragment was reamplified with the same specific primer pair (1G). The amplicon was isolated directly from PCR product using column purification kit (Illustra™).

The expected size of amplified fragment was 325 bp or 224 bp for an amplicon without alpha satellite. Experimentally, we got a short fragment of 169 bp (Figure 8).

Based on sequence analysis, the short fragment seems to originate by two deletion, one of which involved alpha satellite (Figures 8, 9). Alpha satellite element is in red. *Alu*I recognition site is underlined (AGCT).

A

ACAGGTGTGAGCCACTATTCCCCGCCCTTTTTTTTTTTCATTTTTCTACACATTTTA
CTTGTTTGTATTTCTATAGTCTTCTAATGTAGTCTTGAAATTATAAAGGGCTGGGC
ATCGGCTCAAGGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG
AGGCTGAGGTGGGCGGATCAGGAGGTCAGATAAAAAACTAGAAAGA***AGCT***TTCTG
AGAAACTGCTTTGTGATGTGTGCATTCATCTCACAGAGATAAATCTTTCTTTTGAT
TCAGCAGTTTGGAAACCTGGCTAACATGGTGAAACCCGGTGTCTACTA

B

ACAGGTGTGAGCCACTATTCCCAGTTGCTTTTTTTAAAAAAGTTTTTTTGGGTCGG
GCCGGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTAGAT
CACGAGGTCAGGAGTTCGATACCAGCCTGGCCAACATGGTGAAACCCGGTGTCTA
CTA

Figure 8. DNA sequences of *ZNF675* amplicon. A. Sequence (325 bp) generated from NCBI database containing an alpha satellite element (101 bp). B. Sequence of a shorter fragment (169 bp) extracted from agarose gel and sequenced by Sanger method. Alpha satellite element is almost completly deleted.

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)

19.2GenBank                     ACAGGTGTGAGCCACTATTCCCCG-CCCTTTTTTTTTTTCATTTTTCTACACATTTTACT
long19. 2f+R_2017-05-18_004.ab1 ACAGGTGTGAGCCACTATTCCCCGCCCCTTTTTTTTTTTCATTTTTCTACACATTTTACT
short19.2f+R_2017-05-18_004.ab1 ACAGGTGTGAGCCACTATTCCCAG------------TTGCTTTTTTTAAAAAAGTTTTTT
                                ********************** *           ** * *****  * * * ***  *

19.2GenBank                     TGTTTGTATTTCTATAGTCTTCTAATGTAGTCTTGAAATTATAAAGGGCTGGGCATCGGC
long19. 2f+R_2017-05-18_004.ab1 TGTTTGTATTTCTATAGTCTTCTAATGTAGTCTTGAAATTATAAAGGGCTGGGCATCGGC
short19.2f+R_2017-05-18_004.ab1 T------------------------------------------------------------
                                *

19.2GenBank                     TCAAGGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGG
long19. 2f+R_2017-05-18_004.ab1 TCAAGGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGG
short19.2f+R_2017-05-18_004.ab1 ----GGGTCGGGCCGGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGG
                                    *** * ***  ************************************ *******

19.2GenBank                     GCGGATCAGGAGGTCAGATAAAAACTAGAAAGAAGCTTTCTGAGAAACTGCTTTGTGATG
long19. 2f+R_2017-05-18_004.ab1 GCGGATCAGGAGGTCAGATAAAAACTAGAAAGAAGCTTTCTGAGAAACTGCTTTGTGATG
short19.2f+R_2017-05-18_004.ab1 GTAGATCACGAGGTCAG-----------------------------------------
                                *  ***** ********

19.2GenBank                     TGTGCATTCATCTCACAGAGATAAATCTTTCTTTTGATTCAGCAGTTTGGAAACCTGGCT
long19. 2f+R_2017-05-18_004.ab1 TGTGCATTCATCTCACAGAGATAAATCTTTCTTTTGATTCAGCAGTTTGGAAACCTGGCT
short19.2f+R_2017-05-18_004.ab1 -----------------GAG------------TTCGATAC----------CAGCCTGGCC
                                               ***            ** *** *           * ******

19.2GenBank                     AACATGGTGAAACCCGGTGTCTACTA
long19. 2f+R_2017-05-18_004.ab1 AACATGGTGAAACCCGGTGTCTACTA
short19.2f+R_2017-05-18_004.ab1 AACATGGTGAAACCCGGTGTCTACTA
                                **************************
```

Figure 9. CLUSTAL multiple sequence alignment by MUSCLE algorythm. DNA sequences of a long isolated amplicon and short isolated amplicon of element 29 are aligned with the referent sequence from NCBI GenBank database. The results revealed a deletion of alpha satellite element (red) in analysed shorter amplicon of element 29.

3.4. Analysis of DNA methylation at alpha repeats: influence of heat-shock treatment

Genomic DNA was isolated from human fibroblast cell culture at standard conditions and after the heat shock treatment (previously described in Methods). The isolated DNA was quantified by Qubit (Table 11). The quality of isolated DNA was examined by gel electrophoresis (Figure 10).

Table 11. The concentration of genomic DNA isolated from hTERT-MJ90 cell line after the heat-shock treatment quantified by Qubit™.

Treatment	c (DNA) / ng/μL
37 ˚C 2 h	15,00
42 ˚C 2 h	0,96
42 ˚C 2 h	4,39
42 ˚C 2 h, 37 ˚C 1 h	7,31

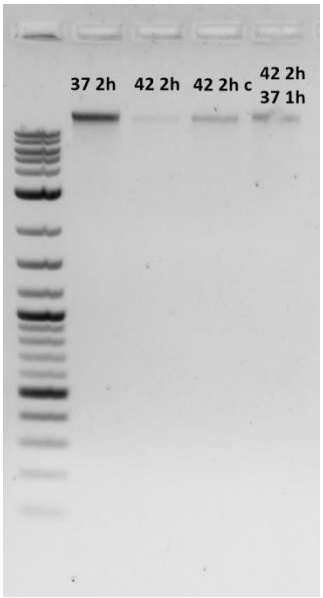


Figure 10. DNA isolated from hTERT MJ90 cells after heat shock treatment. Clear unique bands at the very top of electrophoretic gel indicate intact, undamaged genomic DNA.

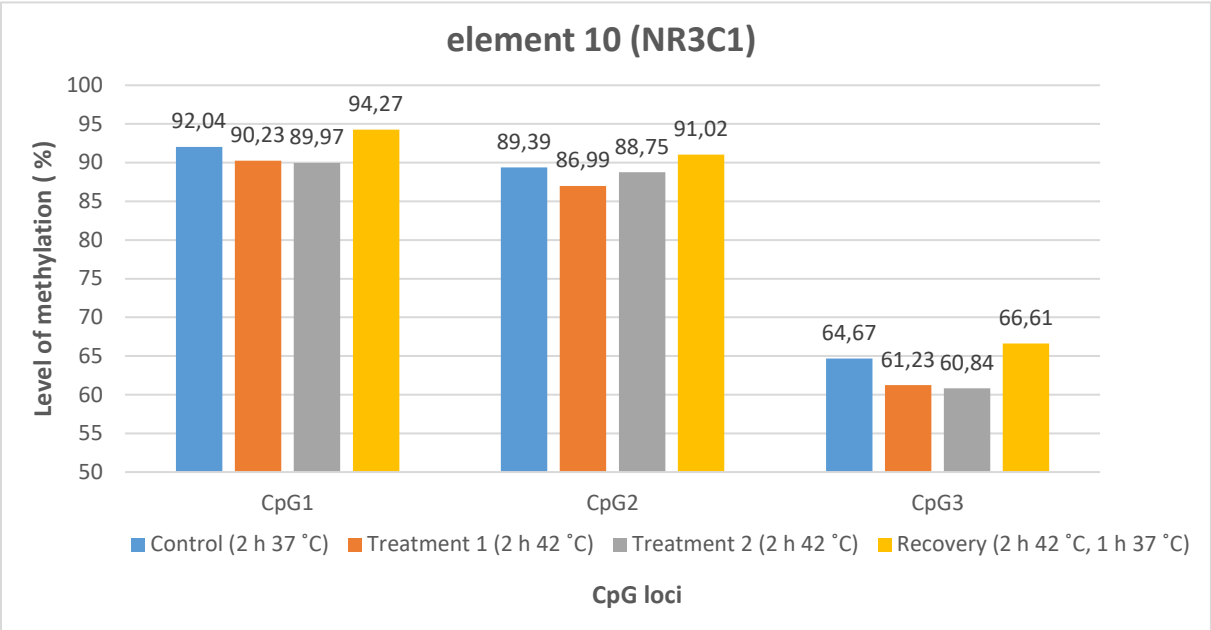
Before the bisulfite treatment, DNA concentrations were uniformed either by diluting or concentrating of DNA solutions to 100 ng. After the bisulfite treatment, fragments for pyrosequencing were amplified using special primers for converted DNA. To enable the attachement of a template DNA strand, reverse primers were biotinylated on their 5'- end.

Target of amplification were the elements in the introns of genes *NR3C1* (Figure 11, element 10) and *ZNF675* (Figure 13, element 29). In total, element 10 with its adjanced regions contains 9 CpG loci, while element 29 contains 4 CpG loci. However, it is important to stress out that the majority of the CpG loci are located in adjanced regions of alpha satellite elements. For instance, only loci number 4, 5, 6 and 7 are located within alpha satellite element 10, while element 29 does not contain a single CpG locus within its sequence.

To obtain the good results, pyrosequencing of element 10 was carried out in two runs, meaning that three CpG loci were analysed in first run and other five in the second one. Four loci of element 29 were analysed in a single run. Since each reaction was performed in duplicates, the methylation status was determined as the average of two measured percentages. The results are processed and graphycally represented in Microsoft Excel (Figure 12, Figure 14).

Figure 11. Bisulfite converted DNA sequence of element 10 (red) in intron of gene *NR3C1* and its flanking regions. CpG loci (highlighted in yellow) analysed by pyrosequencing are numerated 1 to 9. Loci 4, 5, 6 and 7 are located within the element 10 while remaining loci are located in adjanced regions. The underlined sequences represent forward and reverse PCR primers. Pyrosequencing primers are highlighted in blue.

A



B

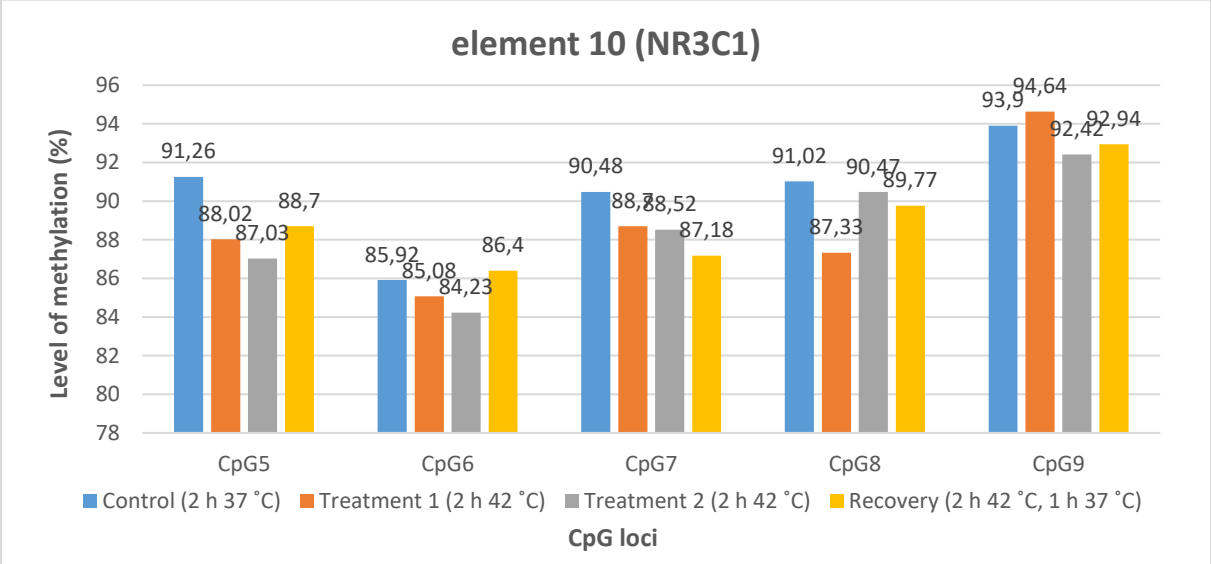


Figure 12. Analysis of DNA methylation level at seven CpG loci for element 10 in intron of gene *NR3C1*. The pyrosequencing was carried out in two runs (A, B). All loci resamble high level of methylation, except CpG3 (A) where methylation does not exceed 66,61 %.

ATAGGTGTGAGTTATTATTTTTCGTTTTTTTTTTTT

TTTATTTTTTTATATATTTTATTTGTTTGTATTTTT

ATAGTTTTTTAATGTAGTTTTGAAATTATAAAGGG

1 2

TTGGGTATCGGTTTAAGGGTTAGGTACGGTGGTT

3

TACGTTTGTAATTTTAGTATTTTGGGAGGTTGAGG

4

TGGGCGGATTAGGAGGTTAGATAAAAATTAGAA

AGAAGTTTTTTGAGAAATTGTTTTGTGATGTGTGT

ATTTATTTTATAGAGATAAATTTTTTTTTTGATTT

AGTAGTTTGGAAATTTGGTTAATATGGTGAAATT

CGGTGTTTATTA

Figure 13. Bisulfite converted DNA sequence of element 29 (red) in intron of gene *ZNF675* and its flanking regions. CpG loci (highlighted in yellow) analysed by pyrosequencing are numerated 1 to 4. All CpG loci are found in adjanced regions. The underlined sequences represent forward and reverse PCR primers. Pyrosequencing primer is highlighted in blue.

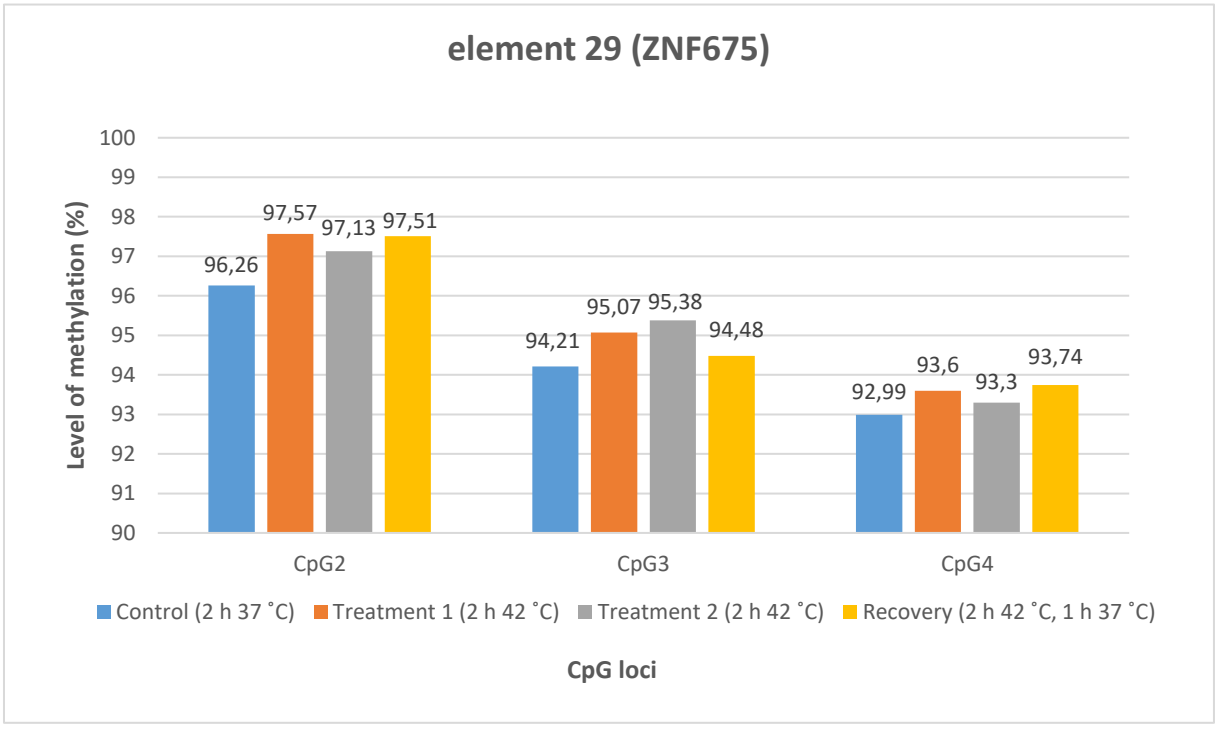
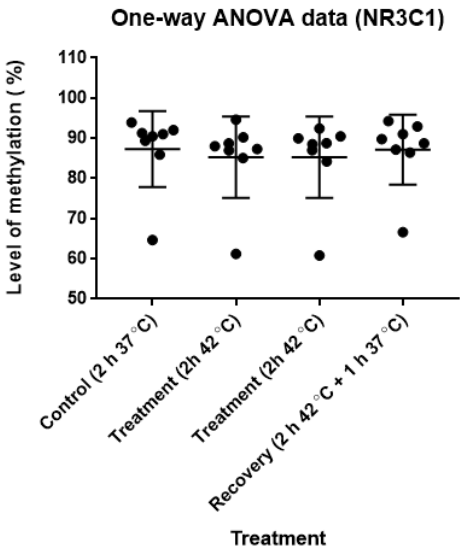


Figure 14. Analysis of DNA methylation at three CpG loci (CpG2, CpG3 and CpG4) for *ZNF675* element 29. All loci resemble high level of methylation.

The results of pyrosequencing revealed high methylation of 8 CpG loci associated with element 10 in intron of gene *NR3C1* and 3 CpG loci associated with element 29 in intron of gene *ZNF675*. The pyrosequencer failed to accurately detect the level of methylation at CpG 4 in *NR3C1* element 10 and CpG 1 in *ZNF675* element 29. For element 10, all CpG loci except CpG 9 showed higher methylation under control conditions (37 ˚C) than under heat stress. The level of methylation after the recovery period varied from site to site. Each locus in element 10 showed methylation above 80% except CpG3 where, surprisingly, the level of methylation did not reach even 70% (Figure 13). Element 29 resambles a slightly different pattern as the average methylation level under the stress conditions is shown to be higher than under control and recovery treatment (Figure 14).

To statisticaly test the significance of the difference between the results, One-way ANOVA with hypothesised non-Gauss distribution was performed in GraphPad Prism 7. The distribution of data values is shown in Figure 15. The results of analysis are shown in Figure 16. Large overall P value implies that data do not exhibit any significant difference between mean values. To consider the difference significant, the overall P value needs to be bellow significant P value (< 0.05). Since the P value = 0.9006 for *NR3C1* and P value = 0.5486 for *ZNF675*, the hypothesis cannot be rejected, meaning that the difference observed between different groups of samples is not statistically significant.

A



B

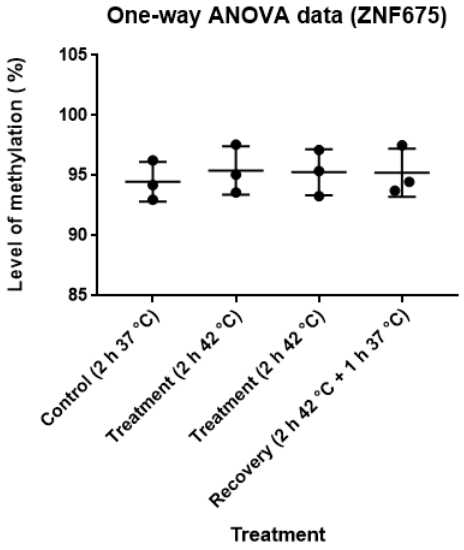


Figure 15. The distribution of ONE-way ANOVA data values for NR3C1 element 10 (A) and *ZNF675* element 29 (B). Data values (y-axis) are grouped by treatment (x-axis). Means and standard deviations for each treatment group are designated.

A

	1way ANOVA	
1	Table Analyzed	One-way ANOVA data
2		
3	Kruskal-Wallis test	
4	P value	0.9006
5	Exact or approximate P value?	Exact
6	P value summary	ns
7	Do the medians vary signif. (P < 0.05)?	No
8	Number of groups	4
9	Kruskal-Wallis statistic	0.7436
10		
11	Data summary	
12	Number of treatments (columns)	4
13	Number of values (total)	12

B

	1way ANOVA	
1	Table Analyzed	One-way ANOVA data
2		
3	Kruskal-Wallis test	
4	P value	0.5486
5	Exact or approximate P value?	Approximate
6	P value summary	ns
7	Do the medians vary signif. (P < 0.05)?	No
8	Number of groups	4
9	Kruskal-Wallis statistic	2.117
10		
11	Data summary	
12	Number of treatments (columns)	4
13	Number of values (total)	32

Figure 16. ONE-way ANOVA with assumed non-Gauss distribution (Kruskal-Wallis test). P values for *NR3C1* element 10 (A) and *ZNF675* element 29 (B) indicate that medians do not vary significantly.

4. Discussion

In this study, for the first time we charachterized dispersed elements of alpha satellite DNA in human genome. We analysed number, size, abundance and distribution of dispersed elements. NCBI database research revealed 31 element of alpha satellite DNA in euchromatin, out of which 10 elements was found in introns of genes and 21 elements in intergenic regions. This pattern of organization has been previously described in red flour beetle *Tribolium castaneum* after whole genome sequencing project, which revealed the association of TCAST satellite elements with 101 protein-coding gene (Brajković et al, 2012).

Of total 31 element, we analysed the presence of 20 elements in genomic DNA isolated from samples of 28 various cell lines and peripheral blood sampled from 26 patients. The localization of 10 elements in introns and 10 elements in intergenic regions is confirmed by PCR. The analysis revealed the difference in abundance of amplicons within population along with the differences between examined cell lines. In most samples, dispersed elements are present at the predicted position in human genome. Experimentaly, this is visible as a single band in the agarose electrophoretic gel with a molecular size that resambles the referent size of a target containing alpha satelite element. In addition to this expected pattern, it has been shown that two distinct amplicons for one locus can be present within one sample, or, in this case, one human individual. Particularly, the amplification of a target alpha satellite element 29 located in intron of gene *ZNF675* resulted with two products that are fully separated in the agarose gel. The amplified fragments experimentally differ in size which approximately equals the size of a hypothesised alpha satellite element, which was subsequentialy confirmed by DNA sequencing and the sequence alignment. However, this was observed only in three samples (R4, R6 and R14) whereas the rest of the examined samples show either single amplicon or no amplicon at all.

Therefore, alpha satellite element 29 seems to be differentially abundant in the population, thus implying the existence of sequence variants that can be considered as polymorphisms. Since either one or both variants can be present in a human individual, we assume that alpha satellite elements are inherited like any other alele – one from mother and one from father. To understand the association between polymorphism and a particular phenotype or behavioural outcome, it is essential to know whether that polymorphism is functional. Most common polymorphisms are potential regulatory polymorphisms located in noncoding regions, such as promoters, upstream/downstream regulatory regions, untranslated regions, introns and intergenic regions. Individual SNPs may have a minimal functional impact

but may be in linkage disequilibrium with a set of polymorphisms that form a haplotype associated with a functional outcome on gene expression or function. For most polymorphisms located in noncoding regions, a variety of approaches, including bioinformatic search for location in consensus DNA elements, can be used to eludicate their function in regulation of transcription. Depending on their location within the noncoding portion of genome, regulatory polymorphisms may affect transcription, RNA splicing, stability or translation, with an exeption of intergenic polymorphisms whose function is yet unknown.

Superhelical tertiary structure of alpha satellite DNA and associated proteins is important for tight packing of DNA into heterochromatin (Ugarković, 2005). The assembly of heterochromatin is regulated through various epigenetic mechanisms, including DNA methylation, which tends to prevent the translocation of repetitive elements. Our special interest was to see if dispersed elements of alpha satellite DNA play the role in structuring of euchromatin and how this process is regulated under different physiological conditions. By determining the methylation status of target regions, we wanted to examine the difference in regulation under the heat stress. Unmethylated DNA adopts an open, accessible conformation that permits interaction with nonchromatin proteins and recruitment of transcriptional machinery. In contrast, methylation of DNA results in a chromatin structure that may prevent transcription factors from accessing DNA or the binding of a transcription factor to its normal recognition site is inhibited by the presence of the methyl group. Also, some specific transcription repressors can directly bind to methylated DNA. (Li, 2002).

Methylation of satellite DNA elements was previously described in red flour beetle *Tribolium castaneum* (Feliciello et al, 2013). The analysis of adult DNA revealed partial presence of 5-mC and/or 5-hmC at 32 out of 33 total cytosine positions with no methylation detected at a single CpT position. Almost the same methylation profile was obtained for embryo DNA, with methylation not being restricted to CpG but is also found at CpA, CpT and CpC sites. Preservation of satellite DNA methylation in all developmental stages of *T. castaneum* indicates the importance of this epigenetic modification for the maintenance of low transcription levels and a silent heterochromatin state. However, the results also indicate the progressive cytosine demethylation after heat-shock, resulting in complete demethylation after prolonged treatment. Such presence of an active DNA methylation/demethylation cycle in *T. castaneum* suggests DNA methylation as an epigenetic mechanism involved in process of environmental adaptation.

To analyse the methylation and the behaviour of alpha satellite elements in response to stress, we carried out a DNA methylation analysis by pyrosequencing after expossing human fibroblast cells to the physiologicaly normal temperature (37 ˚C) and elevated, physiologicaly dangerous temperature (42 ˚C). The DNA methylation analysis after heat shock treatment revealed extremly high methylation of CpG loci. The results as well exhibited a small difference in the percentage of methylation between treated and untreated cells or recovered cells, which happens to be statistically unsupported and insignificant. These findings differ a lot from what is expected and noted in study of DNA methylation in *T. castaneum* (Feliciello et al, 2013) where progressive demethylation in response to heat stress occured.

Based on the given results, we might conclude that the process of heat stress response in human is not regulated through changes in DNA methylation level or that, in fact, the level of methylation is being maintained regardless of stressful conditions. In ectotherm organisms, like insect *T. castaneum* (Feliciello et al, 2013) whose body temperatures conform to ambient temperature, temperature is one of the principal environmental variables that drive adaptive evolution. Consenquently, humans may not exhibit the same regulatory mechanisms regarding the fact they are endotherm organisms. Still, there are few things that need to be taken in consideration.

First of all, we need to take in a count possible errors made throughout the experiment. For example, an inefficient bisulfite conversion would yield many false positive in subsequential pyrosequencing step. Particularly, if unmethylated cytosines were not converted to uraciles and exchanged for thymines, pyrosequencing machine would recognize them as methylated, unconverted cytosines, thus showing the higher methylation. Second, by this method, we can only determine the 5mC in the context of CpG loci, which might not show the full methylation profile. This represented a problem due to a low CG content in alpha satellite elements, so our target regions included CpG loci outside the alpha satellite element. Recent findings show that 5mC can as well be found at non-CpG sites, such as CpA, CpT, and CpC (Jang et al, 2017). We now know that DNA methylation occurs universally at non-CpG sites, which can be a by-product of the hyperactivity of non-specific de novo methylation of CpG sites (Ziller et al, 2011) or it can be correlated with gene expression and tissue specificity (Barrès et al, 2009). Finally, to fully understand the regulational role of dispersed alpha satellite elements, additional experiments are needed. To further investigate the epigenetic mechanisms that might be regulating dispersed elements, a next step would require chromatin-immunoprecipitation (ChIP) analysis which would reveal the histone modifications abundant

under normal and stress conditions. A good follow-up to these experiments would be an examination of nascent siRNAs to spot potential changes in level of their transcription which eventualy could be characterized as a stress response mechanism.

A good experiment to test the regulatory role of dispersed elements of alpha satellite DNA would be the comparisment of heat stress response between two distinct samples regarding the presence of a particular alpha satellite element. The setup would include two group of cells: wild type cells and „knock-out“ cells with the deletion of the alpha satellite element. The deletion can theoretically be obtained by CRISPR/Cas9 technology, but it can be complicated in practice due to the primary charachteristic of target element, and that is its high aboundance in regions of tandemly repeated alpha satellite DNA. Having that in mind, a knock-out of a single dispersed element requires deleting adjanced regions that are not repetitive and are site-specific, meaning that any visible consequence may as well be caused by deletion of adjanced region whose role is unknown or not even hypothesized.

In summary, the exact mechanism of generating and spreading of dispersed elements of alpha satellite DNA across the human genome, as well as the regulatory role of these elements is still not eludicated. Hopefully, future studies will reveal more relevant properties of alpha satellite elements and their impact on associated genes.

5. Conclusion

This study reveals the presence of 31 dispersed element of alpha satellite DNA in euchromatic portion of human genome. Differential abundance of these elements in human individuals indicates they act as polymorphisms. To confirm this phenomenon, further studies need to be carried out on the larger sample of human population. Moreover, the high methylation level of regions associated with alpha satellite elements before and after heat-shock indicates the preservation of methylation profile regardless of stressful condition. Nevertheless, additional studies are needed to confirm the mechanism of distribution of alpha satellite elements across the human genome and to eludicate their role in the structuring of chromatin and modulation of gene expression.

6. References

1. Adams R. L. (1995): Eukaryotic DNA methyltransferases - structure and function. Bioessays 17(2):139-45.
2. Akrap I. (2013): Study of satellite DNA-mediated gene regulation in red flour beetle *Tribolium castaneum* (Herbst, 1797) (graduation thesis). Retrieved from Croatian Digital Theses Repository, National and University Library in Zagreb. https://urn.nsk.hr/urn:nbn:hr:217:628381
3. Barrès R., Osler M. E., Yan J., Rune A., Fritz T., Caidahl K., Krook A., Zierath J. R. (2009): Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. Cell Metabolism 10(3):189-98.
4. Bergmann J. H., Jakubsche J. N., Martins N. M. et al (2012): Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. Journal of Cell Science 125: 411–421.
5. Bewick, V., Cheek, L., Ball, J. (2004). Statistics review 9: One-way analysis of variance. Critical Care, 8(2), 130–136. http://doi.org/10.1186/cc2836
6. Brajković J., Feliciello I., Bruvo-Mađarić B., Ugarković D. (2012): Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*. G3 2: 931-41.
7. Chan F. L., Marshall O. J., Saffery R. et al (2012): Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proceedings of the National Academy of Sciences of the USA 109: 1979–1984.
8. Costello J. F., Plass C. (2001): Methylation matters. Journal of Medical Genetics 38(5):285-303.
9. Davidson E. H., Britten R. J. (1979): Regulation of gene expression: possible role of repetitive sequences. Science 204: 1052-1059
10. Feliciello I, Akrap I, Ugarković Đ (2015): Satellite DNA Modulates Gene Expression in the Beetle *Tribolium castaneum* after Heat Stress. PloS Genetics 11(8): e1005466. doi:10.1371/journal.pgen.1005466
11. Feliciello I., Chinali G., Ugarković Đ. (2011): Structure and population dynamics of the major satellite DNA in the red flour beetle *Tribolium castaneum*. Genetica 139: 999-1008.

12. Feliciello I., Parazajder J., Akrap I., Ugarković Đ. (2013): First evidence of DNA methylation in insect *Tribolium castaneum*. Epigenetics 8(5): 534-541. doi: 10.4161/epi.24507
13. Fraga M. F., Rodríguez R., Cañal M. J. (2000): Rapid quantification of DNA methylation by high performance capillary electrophoresis. Electrophoresis 21(14):2990-4.
14. França L. T., Carrilho E., Kist T. B. (2002): A review of DNA sequencing techniques. Quarterly Reviews of Biophysics 35(2):169-200. Review.
15. Frommer M., McDonald L. E., Millar D. S., Collis C. M., Watt F., Grigg G. W., Molloy P. L., Paul C. L. (1992): A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences U S A 89(5):1827-31.
16. Gonzalgo M. L., Liang G., Spruck C.H., Zingg J. M., Rideout W. M., Jones P. A. (1997): Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. Cancer Research 57(4):594-9.
17. Hall L. E. , Mitchell S. E., O'Neill R. J. (2012): Pericentric and centromeric transcription: a perfect balance required. Chromosome Research 20: 535–546.
18. Howell, David (2002). Statistical Methods for Psychology. Duxbury. pp. 324–325. ISBN 0-534-37770-X.
19. Jang H. S., Shin W. J., Lee J. E., Do, J. T. (2017). CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. Genes, 8(6), 148. http://doi.org/10.3390/genes8060148
20. Li L.-C., Dahiya R. (2002). MethPrimer: designing primers for methylation PCRs. Bioinformatics. Vol 18. 11: 1427-1431
21. Maio J. J., (1971): DNA strand reassociation and polyribonucleotide binding in the African green monkey *Cercopithecus aethiops*. Journal of Molecular Biology 56: 579–595.
22. Nyrén P., Lundin A. (1985): Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Analytical Biochemistry 151(2):504-9.
23. Pezer Ž., Brajković J., Feliciello I., Ugarković Đ. (2012): Satellite DNA-Mediated Effects on Genome Regulation: Genome Dynamics 7: 153-169.

24. Rizzi N., Denegri M., Chiodi I., Corioni M., Valgardsdottir R., Cobianchi F., Riva S., Biamonti G. (2004): Transcriptional activation of a constitutive heterochromatic domain of the human genome inresponse to heat shock. Molecular Biology of the Cell 15: 543-51
25. Ronaghi M., Karamohamed S., Pettersson B., Uhlén M., Nyrén P. (1996): Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry 242(1):84-9.
26. Royo J. L., Hidalgo M., Ruiz A. (2007): Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. Nature Protocols 2: 1734-1739. doi:10.1038/nprot.2007.244
27. Siegel S., Castellan N. J. (1988): Nonparametric Statistics for the Behavioral Sciences (Second ed.). New York: McGraw–Hill. ISBN 0070573573.
28. Stirzaker C., Millar D. S., Paul C. L., Warnecke P. M., Harrison J., Vincent P. C., Frommer M., Clark S. J. (1997): Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. Cancer Research 57(11):2229-37.
29. Ting D. T., Lipson D., Paul S. et al (2011): Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science 331: 593–596.
30. Ugarković Đ. (2005): Functional elements residing within satellite DNAs. EMBO Reports 6: 1035-1039
31. Ugarkovic Đ. (2013): Evolution of Alpha-Satellite DNA. eLS. John Wiley & Sons, Ltd: Chichester. doi: 10.1002/9780470015902.a0020829.pub2
32. Ugarković D., Plohl M. (2002): Variation in satellite DNA profiles--causes and effects. EMBO Journal 21(22):5955-9
33. Ziller M. J., Müller F., Liao J., Zhang Y., Gu H., Bock C., Boyle P., Epstein C. B., Bernstein B. E., Lengauer T., Gnirke A., Meissner A. (2011): Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genetics 7(12):e1002389.
34. Zofall M., Grewal S. I. (2006): RNAi-mediated heterochromatin assembly in fission yeast. Cold Spring Harbour Symposia on Quantitative Biology 71:487-96.
35. https://www.qiagen.com/dk/resources/technologies/pyrosequencing-resource-center/technology-overview/

7. Curriculum vitae

NAME AND SURRNAME: Doris Zorić

DATE AND PLACE OF BIRTH: March 27th, 1994; Zadar, Croatia

EDUCATION

2015 – 2018: **Master studies in molecular biology**; Faculty of Science, University of Zagreb, Croatia; Graduation thesis: Charachterization of dispersed elements of alpha satellite DNA in human genome

2012 – 2015: **Bachelor studies in molecular biology**; Faculty of Science, University of Zagreb, Croatia

2008 – 2012: **High school education**; Franjo Petrić Gymnasium, Zadar, Croatia

RESEARCH EXPERIENCE

12/2016 – 06/2017: **Master student**; Ruđer Bošković Institute, Zagreb, Croatia; Laboratory of Evolutionary Genetics (LEG); Dr. Đurđica Ugarković, Assoc. Prof

29/08/2016 – 14/10/2016: **Research assistant**; Institute for Molecular and Cell Biology (IMCB), Singapore; Microfluidic Systems Biology (MSB) Lab

03/2016 – 06/2016: **Traineeship**; Faculty of Science, University of Zagreb, Croatia; Department of Molecular Biology; Dr. Petra Korać, Asst. Prof.

03/2015 – 02/2016: **Traineeship**; Ruđer Bošković Institute, Zagreb, Croatia; Laboratory for Molecular and Cell Biology; Cell Ageing and Telomere Analytics; Dr. Ivica Rubelj, Assoc. Prof.

10/2013 – 02/2014: **Demonstrator in Zoology**; Faculty of Science, University of Zagreb, Croatia

LANGUAGES

Croatian (mother tongue), **English** (independent user; TOEFL (104/120); B2), **Spanish** (basics), **German** (basics)

HONOURS, AWARDS, GRANTS, FELLOWSHIPS

2018: **ERASMUS + Student Mobility of Placement (SMP)**; Chalmers University of Technology, Goteborg, Sweden; SYSBIO

2016: **Singapore International Pre-Graduate Award (SIPGA)**; Agency for Science, Technology and Research (A*STAR), Singapore

2015: ***Green Power Volume 2: Let's Live Sustainable***; ERASMUS+ Youth Exchange; Mine Vanagnti NGO; Uri, Sardegna, Italy

PROJECTS

2016: ***8th Biology Night***

2014: ***6th Biology Night***

2013: ***EntArach***

2013: ***Apsyrtides***

MEMBERSHIPS

2017 – present: **Croatian Academic Sports Association – SUBOS**

2012 – 2014: **Biology Student Association (BIUS)**